Inhibition and Allosteric Signaling of Transcription Activation by Bacterial AraC Family Activator Proteins

By

Jiaqin Li

Submitted to the graduate degree program in the Department of Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date Defended: Oct 31, 2016

The Dissertation Committee for Jiaqin Li

certifies that this is the approved version of the following dissertation:

Inhibition and Allosteric Signaling of Transcription Activation by Bacterial AraC Family Activator Proteins

_____

Chairperson: Dr. Susan Egan

Date approved: Nov. 22, 2016

Abstract

AraC family transcriptional activators are defined by a 100-amino acid DNA-binding domain (DBD) containing two DNA binding helix-turn-helix (HTH) motifs. My research focused on three AraC family proteins: RhaR (activator of the *E. coli* L-rhamnose catabolic regulon), VirF (activator of expression of the *Shigella flexneri* type three secretion system), and ToxT (activator of *Vibrio cholera* virulence gene).

By using fluorescence-based thermal shift assay, and intrinsic tryptophan fluorescence assay, I have shown that small molecule inhibitor SE-1 directly binds to VirF and RhaS-DBD. Mutagenesis studies of residues in RhaS support the prediction of docking that SE-1 is likely bound to a small pocket between the two HTH motifs. With a final goal to provide direct evidence of the position on the ToxT protein that serves as the SE-1 binding site, ToxT was crystalized under solution conditions different than the previously published ones, and the structure of ToxT was determined to a higher resolution than the published structure. A region that was missing from the previously determined structure now can be traced entirely. Using Electrophoretic Mobility Shift Assay (EMSA), SE-1 was found to inhibit DNA binding by ToxT. Co-crystallization trials of SE-1 and ToxT were carried out. Unfortunately, I was not able to obtain any crystals of ToxT-SE-1.

In collaboration with Dr. Jeff Aubé, I have tested SE-1 analogues with the ultimate goal of optimizing the potency and specificity of SE-1. Unfortunately, no analogs were found with increased VirF inhibition potency relative to SE-1. In the process of synthesizing analogs, we found that SE-1 and its analogues converted to corresponding quaternary salts in aqueous solution, and the quinolinium salt was responsible for the observed inhibition by SE-1.

In this study, I also investigated the mechanism by which RhaR responds to L-rhamnose. During these studies, I discovered that the RhaR start codon was previously annotated 30 codons upstream of the true start codon. The equilibrium binding affinity of RhaR to its full and half DNA binding sites, of the isolated RhaR-DBD to the half-site, and the rhamnose dependence of DNA bending by RhaR were measured. The findings support a model in which RhaR NTD increases the DNA binding affinity of each RhaR protomer DBD, independent of rhamnose; and that rhamnose signalling primarily increases the positive cooperativity of DNA binding by the two DBDs in a RhaR dimer. These findings suggest a model for the mechanism of allosteric rhamnose signalling in RhaR.

To provide more structural information about RhaR protein, the structures of RhaR-NTD were determined in the presence and absence of L-rhamnose. The 2.05Å rhamnose-bound RhaR-NTD structure showed that the protein forms an antiparallel dimer, and shared a fold that was similar to the AraC-NTD, binding its respective sugar L-rhamnose within a β-barrel. In addition, a Ni^{2+} ion, which has not been seen in other AraC family protein structures, was present in the sugar-binding, cupin superfamily, motif of RhaR. A rhamnose-free structure was also solved to 1.73 Å, and in this structure, a loop region that is involved in rhamnose binding was completely disordered. A second loop region also has minor structural changes. Each of the two regions with rhamnose-dependent structural changes is predicted to be at the interface between the RhaR NTD and DNA-binding domain, suggesting their potential involvement in rhamnose allosteric signaling. No differences were observed in the RhaR N-terminal arm region in the fully and partially rhamnose-occupied structures, suggesting that RhaR rhamnose-dependent allosteric signaling shares some features with the 'light switch' model of AraC, but differs in other features.

Acknowledgements

I would like to take this opportunity to acknowledge my mentor Dr. Susan Egan for her continuous guidance throughout my research in her laboratory. All the success I achieved in this laboratory will be truly credited to her. It would have been impossible for me to finish my projects and dissertation without her outstanding supervision. She spent enormous number of hours in training me on research techniques, problem solving and most importantly in improving my presentation skills. I feel short of words to express my gratitude towards her.

With sincere gratitude I acknowledge my committee members Dr. Scott Hefty, Dr. Eric Deeds, Dr. Christian Ray, and Dr. Joanna Slusky for their comments, suggestions and help over the years. I record my sincere thanks to my colleagues and friends Graham, Deena, Veer, Bria, and Jeff for helping me either directly or indirectly during my PhD. In particular, I am thankful to Jeff and Veer for their unconditional help and willingness to answer my questions all the time. I am also grateful to Graham and Deena for their tremendous amount of supports and contributions to my research.

All my achievements would be meaningless if I fail to acknowledge my parents. Your love, affection and constant support has made me what I am today. They understood the challenges I have to face during my graduate career and wholeheartedly supported me with a lot of patience. I am sure you will be proud of my achievements.

Table of Contents

Chapter 1

INTRODUCTION

AraC Family of Transcriptional Activators

The AraC family is a large family of bacterial transcriptional activator proteins that have been identified in over 70% of sequenced bacterial genomes [1]. The defining feature of members of the AraC family is a 100 amino acid stretch comprising a DNA binding domain (DBD) with two helix-turn-helix (HTH) motifs [2-6]; these HTH motifs bind in consecutive major grooves of specific DNA sequences that are unique for each AraC family paralog [7, 8]. Although some family members such as MarA and SoxS consist of a single domain, most AraC family members also have a domain that we refer to as the regulatory domain. The regulatory domain is typically directly responsible for dimerization and/or effector binding, and if protein activity responds to an effector, allosterically regulates DBD activity [3, 4, 8]. A large subset of homologs share sufficient sequence similarity with the regulatory domain of AraC [3, 4] that they are expected to share structural similarity. This subset includes the activators of the *E. coli* L-rhamnose regulon, RhaS and RhaR; VirF, the master regulator of the type three secretion system (T3SS) in *Shigella flexneri*; and ToxT, a transcriptional activator of virulence gene expression in *Vibrio cholera*.

AraC family proteins regulate expression of genes that are involved in carbon metabolism, stress responses, and virulence [1-4, 9]. Proteins that regulate expression of genes involved in carbon metabolism include RhaS and RhaR (catabolism of L-rhamnose in *E.coli*) [10], AraC (catabolism of L-arabinose in *E.coli*) [11], XylR (catabolism of xylose in *E. coli*) [4], XylS (catabolism of benzene derivatives in *Pseudomonas putida*) [12], MelR (catabolism of melibiose in *E.coli*) [13]. The second group of AraC family regulators controls stress responses, for instance Ada (response to alkylating agents in *E. coli, Salmonella typhimurium* and *Bacillus subtilis*) [14-16], SoxS (response to oxidative stress in *E.coli*) [17], MarA (response to antibiotics

in *E. coli*) [18]. The third group of AraC family regulators regulates expression of genes related to virulence, for example ToxT (regulates expression of cholera toxin, the toxin-coregulated pilus, and its own expression in *Vibrio cholerae*) [19-23], Rns (regulates expression of CS1 and CS2 pili from Enterotoxigenic *E. coli*) [24], VirF (master regulator of T3SS gene expression from *Shigella flexneri*) [25], and ExsA (regulates expression of T3SS related genes in *Pseudomonas aeruginosa*) [26].

Mechanism of transcription activation by AraC

AraC is the founding member of the AraC family proteins and responds to its effector arabinose to regulate expression of the L-arabinose regulon using a well-characterized molecular mechanism known as the "light switch" mechanism (Fig. 1) [27-31]. In addition, although no full-length structure is yet available, high-resolution structures are available for both domains of AraC, including structures of the N-terminal regulatory domain (NTD) in the absence and presence of the effector arabinose [32-34]. In the "light switch" model of AraC, residues at the N-terminus of AraC, a region known as the "arm", play an important role in transmitting the arabinose-binding signal from the AraC NTD to its DBD. In the absence of arabinose, the arm is required to constrain the AraC dimer such that it forms a DNA loop between two well-separated protomer-binding (half) DNA sites in its repressing state. When arabinose binds to the AraC NTD, the conformation of the arm changes, releasing the constraint on the AraC dimer such that AraC can to bind to adjacent half-sites and activate transcription from the *araBAD* promoter [35]. Recent studies provide evidence that arabinose binding may affect the flexibility of the AraC inter-domain linker, which in turn could alter the conformation and DNA binding preferences (well separated versus adjacent) of the dimer [36]. Overall, the binding of effector to

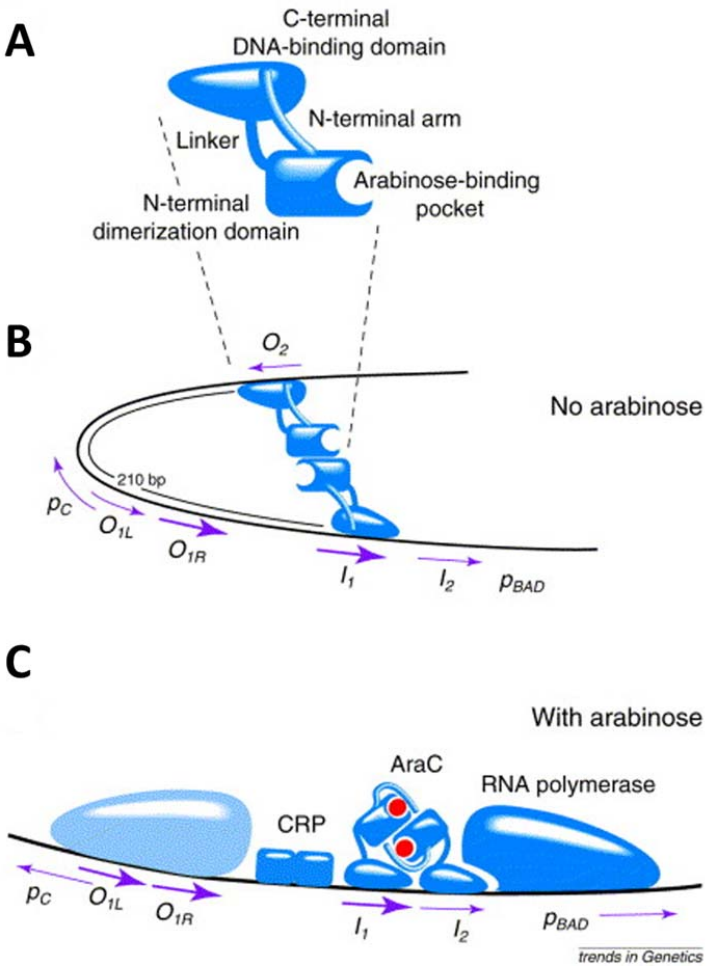


Fig. 1. The AraC "Light switch" Mechanism. (A) The domain structure of one protomer of AraC protein. **(B)** In the absence of the arabinose, the N-terminal arm of AraC contacts the DNA binding Domain (DBD), constraining the conformation of the protein such that one protomer contacts the distal O_2 half-site, resulting in DNA looping and inhibition of *araBAD* transcription. **(C).** In the presence of arabinose, the arm folds over the arabinose-binding pocket, allowing the protein to contact the adjacent I_1 and I_2 half-sites, thus activating transcription. From Schleif (2000) [37].

AraC is hypothesized to reposition the two DBDs in a dimer relative to the N-terminal effector-binding domains, which for AraC results in an increased affinity for adjacent half sites. Most characterized AraC family proteins do not use DNA looping as part of their allosteric signalling mechanisms. However, analogous effector-dependent rigid body movement of DBDs could, in principle, toggle proteins between states with greater (activating state) or lesser (non-activating state) capability to activate transcription; with transcription activation increasing with DNA binding affinity and/or ability to contact RNA polymerase.

Mechanism of transcription activation by ToxT

ToxT is an AraC family transcriptional activator of *Vibrio cholera* virulence gene expression with a C-terminal DNA binding domain and an N-terminal domain involved in dimerization and effector binding [38]. ToxT directly activates the expression of the genes encoding the toxin-coregulated pilus, which is essential for colonization of the human intestine, and the cholera toxin, the cause of the diarrheal disease that is characteristic of cholera [19-21]. ToxT has also been shown to positively auto-regulate its own expression from the *tcp* promoter [22, 23]. In *V. cholera*, ToxT-dependent gene activation was inhibited by both bile and individual unsaturated fatty acids found in bile, such as arachidonic acid, linoleic acid, oleic acid, and plamitoleic acid (PAM) [39-41]. Unlike AraC, ToxT is in its non-activating state in the presence of its effector [38, 40]. The crystal structure of the full-length ToxT protein in its non-activating state with effector bound has been solved [38] (Fig. 2a). A model for the mechanism of transcription activation by ToxT has been proposed based on structural and genetic analysis [38, 41] (Fig. 2b). According to this model, the presence of the effector fatty acid keeps ToxT in a “closed” conformation that is unable to dimerize and not favorable to bind DNA by bridging

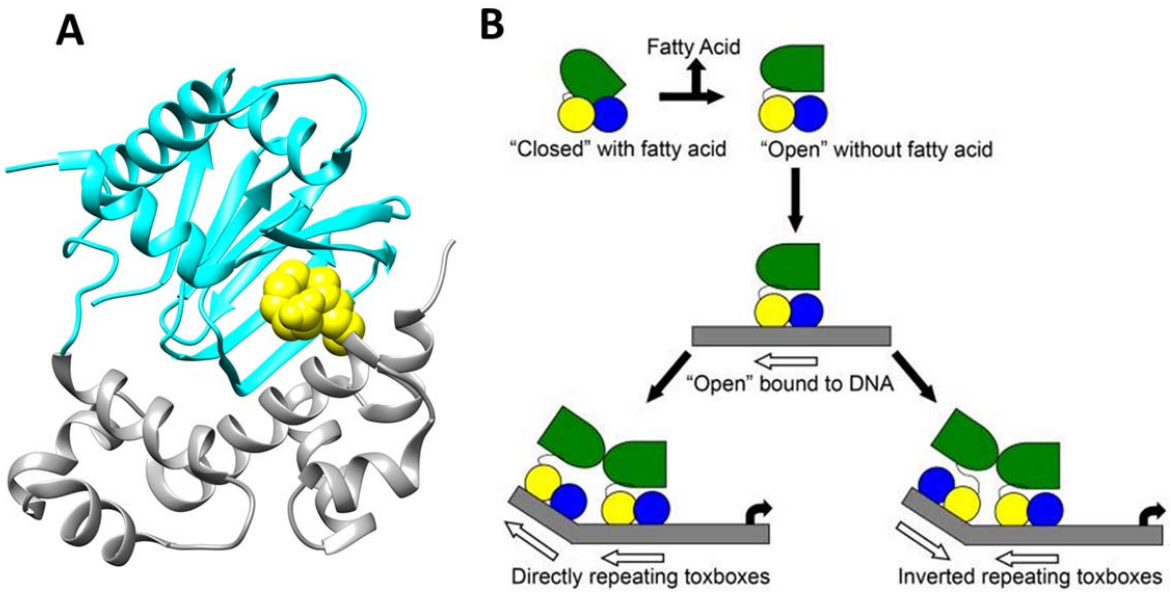


Fig. 2. The Structure and model for the regulation of ToxT. **(A)** Structure of ToxT (PDB code: 3GBG) [38]. Ribbon diagram of the ToxT with the effector palmitoleic acid (PAM) bound in the NTD. The N-terminal domain is in cyan, the DNA binding domain is in grey, the PAM is in yellow sphere. **(B)** Model for the regulation of ToxT by monounsaturated fatty acids. ToxT is in a "closed" conformation in the presence of fatty acid, which cannot bind to DNA. In the absence of fatty acid, ToxT is in an "open" conformation that can bind to DNA. In the "open" conformation the N-terminal domain is free to move in relation to the C-terminal domain, and is able to dimerize with another ToxT at an adjacent toxbox. The linker is sufficiently flexible to allow dimerization on DNA in either direct or inverted orientations. From Lowden *et al* (2010) [38].

residue Lys31 from NTD and Lys230 from DBD through the carboxylate head group of the fatty acid. The inter-domain interaction mediated by effector in this "closed" form, is likely a mechanism for transmitting the effector binding signal from the NTD to the DBD [38]. In the absence of the effector, Lys31 and Lys230 no longer can be bridged by the effector, therefore the inter-domain interaction is lost and the protein is converted to an "open conformation" which dimerizes and is capable of binding DNA and activating transcription [38]. Alternatively, given the large ToxT inter-domain interface (~2000 $Å^2$)[38], and the presence of inter-domain contacts in AraC in both its (+) and (-) arabinose states [42], we propose that the ToxT inter-domain contacts may be altered and/or rearranged, but not eliminated, in its activating state.

The L-Rhamnose Regulon

The L-rhamnose regulon of *E. coli* consists of six genes that are grouped into three operons: *rhaSR*, *rhaBAD* and *rhaT* (Fig. 3) [10, 43-46]. The *rhaSR* operon encodes the activators RhaS and RhaR, which are both AraC family members [2, 4, 6, 46]. The *rhaBAD* operon encodes three enzymes required for L-rhamnose catabolism - rhamnulokinase, L-rhamnose isomerase and rhamnulose-1-phosphate aldolase, respectively [44, 47]. The *rhaT* operon encodes the L-rhamnose -proton symporter that transports L-rhamnose into the bacterial cell [45]. RhaS activates transcription of both the *rhaBAD* and the *rhaT* operons in response to L-rhamnose [10, 45], and RhaR activates the expression of the *rhaSR* operon, also in the presence of L-rhamnose. RhaS and RhaR activate transcription to maximal levels (above basal levels) only in the presence of their common effector, L-rhamnose. In the absence of L-rhamnose, all three of the L-rhamnose-regulated operons in *E. coli* are expressed at basal levels [10, 48, 49]. When L-rhamnose becomes available, basal levels of RhaR protein activate expression of the *rhaSR* operon [10, 46], and the increased level of RhaS protein then activates

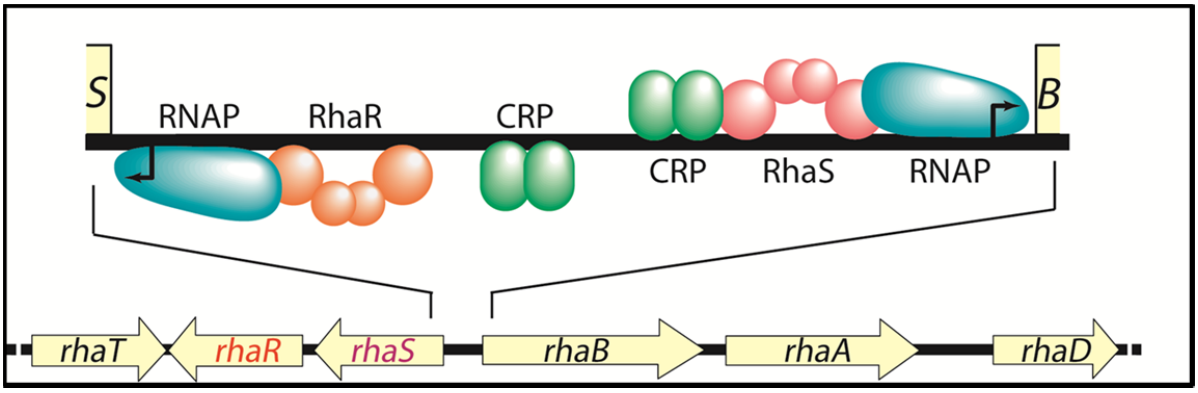


Fig. 3. Schematic Representation of the *Escherichia coli* L-rhamnose Regulon. Thick horizontal black lines represent DNA. Top: Expanded view of the regulatory region between the *rhaSR* and *rhaBAD* operons with the approximate positions of the known proteins bound to DNA at *rhaBAD* and *rhaSR*. Bent arrows indicate transcription start sites. RhaR is in orange, RhaS in pink, CRP in green, and RNAP in teal. Letters "S" and "B" at the ends of the top line represent the beginning of the *rhaS* and *rhaB* genes, respectively. Bottom: Representation of the orientation and direction of transcription for the genes in the L-rhamnose regulon. Courtesy of Dr. Susan Egan.

transcription of the *rhaBAD* and *rhaT* operons [10, 44, 45, 50]. The full transcription activation of all three operons within the L-rhamnose regulon requires a second protein, CRP [10, 49, 51]. CRP binds upstream of the RhaS and RhaR DNA binding sites, at positions centered at -92.5 and -93.5 at *rhaBAD* and *rhaT* [10, 49], respectively, and at -111.5 at *rhaSR* [10].

Transcriptional Activators RhaS and RhaR

RhaS and RhaR are predicted to have similar structures due to their relatively high levels of sequence identity (35%) and similarity (67%) with each other. RhaS is 278 amino acids in length and RhaR is longer with 312 amino acids [46]. Previous study has shown that the extra length of RhaR is mostly due to a 33 amino acid extension at the N-terminal end. The function of this 33 amino acid extension is unknown since deletion of the extension did not change transcription activation *in vivo* [52]. Both RhaS and RhaR function as homodimers, and contain two domains. One domain is a C-terminal, AraC family DNA-binding domain (DBD). The DBDs bind two 17-bp inverted repeat DNA half sites, which are separated by 16 bp (for RhaS) or 17 bp (for RhaR), and contacts RNA polymerase[50, 51, 53, 54]. The other domain is an N-terminal regulatory domain (NTD) that contains an L-rhamnose binding site and a dimerization interface [50, 53, 55].

In the absence of L-rhamnose, RhaS is in a conformation that is unfavorable to bind DNA; however RhaS DBD can bind DNA and activate transcription [10, 55]. This suggests that the NTD in the absence of L-rhamnose limits (inhibits) RhaS binding to DNA; however in the presence of L-rhamnose, inhibition of DNA binding is relieved through allosteric signaling. In contrast, although both RhaR and RhaR DBD can detectably bind DNA in the absence of L-rhamnose, only full-length RhaR can fully activate transcription when L-rhamnose is present [48, 53, 55]. This suggests that the RhaR NTD in the absence of L-rhamnose does not inhibit the

binding of RhaR to DNA. Previous mutagenesis study showed that RhaS and RhaR both make protein-protein interactions with the σ^{70} subunit of RNA polymerase to activate transcription [54, 56].

Previous work leads us to propose a model that the binding of L-rhamnose to RhaR increases both DNA binding and transcription activation [48, 53, 55], and this increase is likely due to structural and/or dynamics changes in RhaR upon L-rhamnose binding; the signal of L-rhamnose binding must be transmitted from NTD to DBD of RhaR to toggle RhaR between its non-activating and activating states. Previously, our lab has constructed single alanine substitutions in the RhaR linker region connecting the NTD and DBD, and found that all the variants activated transcription to the same extent as the wild-type RhaR in the presence of L-rhamnose and no variants has shown an increase in transcription activation over 1.5-fold [57]. The results suggested that none of the residues tested in the linker region is crucial for the L-rhamnose response or transcription activation, thus may not play an important role in activation [57]. However, recent study of AraC protein indicated that the linker region is possibly transition from a unstructured loop to α-helix in response to L-arabinose [36]. Recently, the structure of the regulatory domain of ExsA, an AraC family protein from *P. aeruginosa*, was solved, and the structure was expected to represent the activating conformation of ExsA [58]. Analysis of the ExsA regulatory domain revealed that the linker region is in a helical conformation instead of unstructured. Therefore, it is likely that the allosteric signal that converts RhaR from its non-activating state to its activating state involves L-rhamnose-dependent changes in the inter-domain contacts, and/or structural change in the linker region.

AraC family virulence regulator VirF

VirF is an AraC family transcriptional activator from *Shigella*, the causative agent of bacillary dysentery, or shigellosis [59]. Various pathogenic species of *Shigella*, such as *S. flexneri*, *S. dysenteriae, S. sonnei and S. boydii* are responsible for 165 million illnesses and over 1.1 million deaths worldwide each year, with 70% of the deaths occurring in children under the age of five [60, 61]. All of these species harbor a 230 kb virulence plasmid, of which a 31 kb region named the "entry region" encodes proteins required for the invasion of the bacteria into colonic and rectal epithelial cells, the formation of the type 3 secretion system (T3SS) machinery, invasion of *Shigella* into host epithelial cells, and cell-to-cell spread [62-66].

The regulation of *Shigella* virulence gene expression is largely controlled by the activity of VirF in response to many environmental signals, including temperature, osmolality, and pH [59, 67-69]. The primary regulator VirF positively activates the transcription of two virulence-associated genes, *virB* and *icsA* (Fig. 4). VirB protein then activates transcription of several virulence-associated genes, including *icsB* which is important for intercellular spread [70], *ipaB* which is required for epithelial cell entry and phagosome escape [64], and *mixE* which regulates the expression of many other virulence-associated genes [71]. IcsA encodes a *Shigella* adhesin that facilitates intracellular spread through actin-based motility [72, 73]. Given that VirF controls the regulation and gene expression of the entire T3SS, any factor that could influence the activity of VirF would affect the expression of *virB* and *icsA*, and consequently impact *Shigella* virulence. Therefore, small molecules that target VirF have the potential to be developed into new antimicrobial agents.

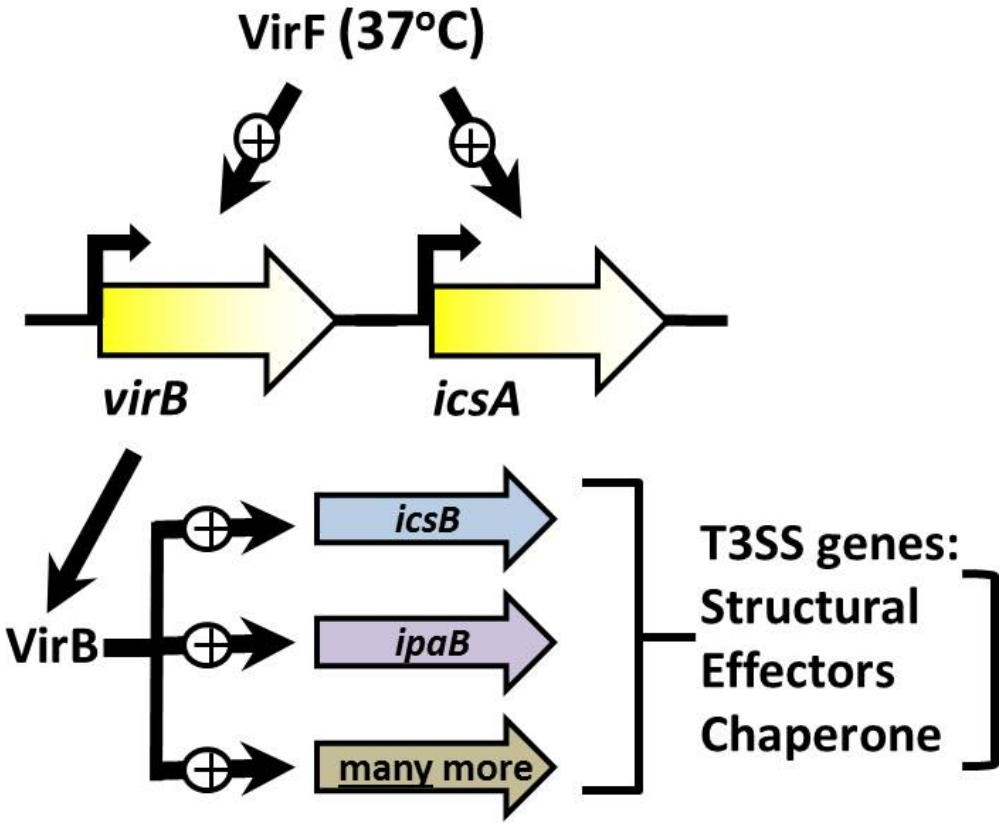


Fig. 4. VirF activates transcription of the *virB* and *icsA* operons [69]. VirB in turn activates expression of genes in *mxi*, *spa* and *ipa* operons that are required for the formation of T3SS machinery, invasion and cell-cell spread of *Shigella* into host epithelial cells [62-66]. (Figure courtesy of Dr. Susan Egan).

Small molecule inhibitor SE-1

Antibiotic resistance has become a worldwide problem that threatens the treatment and prevention of bacterial infections. Each year in the United States, at least 2 million people infected with antibiotic-resistant bacteria, and at least 23,000 people die each year as a direct result of these antibiotic-resistance infections [74]. Traditional antibiotics target bacterial processes essential for the growth, including cell wall synthesis, protein synthesis, DNA replication, and RNA transcription. One alternative strategy that targets virulence factors has been reported in recent literature [75-82]. As mentioned earlier, members of the AraC family proteins regulate bacterial virulence and stress response [1-4, 9], which makes them ideal targets for this unique strategy. Since a structurally conserved DNA-binding domain is present in all of the AraC family proteins, it is expected that a small molecule inhibitor affecting this domain would impact the activity of the proteins and thus act as an effective microbial agent for preventing diseases. In fact, several studies have identified small molecule inhibitors that target specific AraC family proteins [76-79, 82-84].

Using high-throughput screening, our lab has identified a small molecule compound, SE-1 (previously called OSSL_051168) (Fig.5), as an inhibitor of the AraC family protein RhaS [81]. Further study has also shown that SE-1 inhibited several other AraC family regulators, including RhaR, ToxT, and VirF [80, 81]; however, SE-1 does not block DNA binding by the non-AraC family proteins LacI and CRP [81], which suggested that SE-1 has specificity towards AraC family proteins. It was found that SE-1 inhibited DNA binding by the DNA-binding-domain (DBD) of RhaS to the same extent as the full-length RhaS, which indicated that the DBD likely was the target of SE-1 binding [81]. SE-1 not only has been found to inhibit the DNA

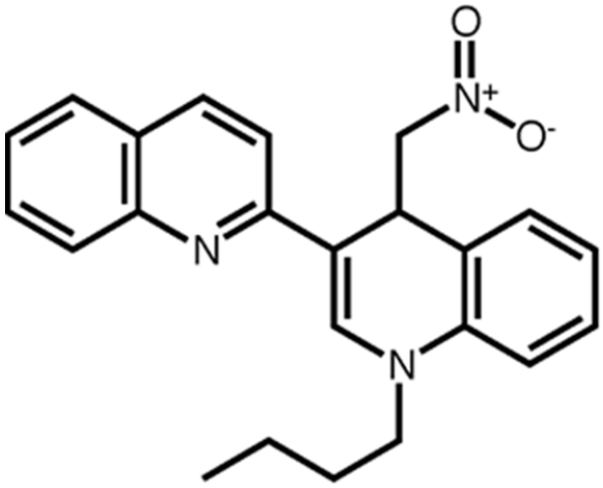


Fig. 5. Structure of SE-1. Chemical structure of SE-1, 1-butyl-4-nitromethyl-3-quinolin-2-yl-4*H*-quinoline. From [81].

binding by VirF, it also has been demonstrated that SE-1 reduced the expression of many virulence genes regulated by VirF, including *icsA*, *virB*, *icsB*, and *ipaB* in *Shigella*, and the invasion to the host cell by *Shigella* [80]. Therefore, SE-1 has the potential to be developed into a novel antibacterial agent since it exhibits selectivity toward AraC family protein.

Research goals

The overall goals of my research are to identify small molecule inhibitor SE-1 binding site of AraC family regulators, to optimize SE-1, and to better understand the transcriptional activation mechanisms by RhaR. The overall goals can be divided into the following sub-goals.

Identify the binding site of the small molecule inhibitor SE-1 on AraC family regulators. The first goal of my research was to test whether SE-1 directly binds to purified VirF and RhaS proteins, and identify where SE-1 may bind to the protein RhaS. Using fluorescence-based thermal shift assay, I found that the addition of 80 μM SE-1 increased the T_m of purified VirF protein (1 μM) by 0.61°C (±0.09), which support the hypothesis that SE-1 directly binds to VirF. Using an intrinsic tryptophan fluorescence assay, I found that there was a concentration dependent quenching of purified RhaS DBD protein fluorescence upon addition of SE-1 that saturated at ~7% reduction in the fluorescence peak signal. This result supports the hypothesis that SE-1 directly binds to RhaS, specifically RhaS DBD. Mutagenesis study of residues in RhaS that were predicted in the SE-1 binding pocket by computational docking suggested that residues D191, S249, and L257 were at or near the SE-1 binding site on RhaS-DBD, which supports the prediction of docking that SE-1 is likely bound to a small pocket between the two HTH motifs.

Crystallization of of ToxT from *Vibrio cholera.* The second goal of my research was to test whether SE-1 inhibited *in vitro* DNA binding by ToxT, and set up crystallization trials to co-crystalize SE-1 with ToxT to provide direct evidence of the position on the ToxT protein that serves as the SE-1 binding site. Using Electrophoretic Mobility Shift Assay (EMSA), I found that SE-1 inhibited DNA binding by ToxT. I further crystallized ToxT using solution conditions that were different than the published conditions, and solved the structure of ToxT to a higher resolution than the published structure. A region that was missing from the previously determined structure now can be traced entirely. Then I started to crystalized SE-1 and ToxT. Unfortunately, I wasn't able to obtain any crystals of ToxT-SE-1 due to many reasons.

SE-1 Optimization. The third goal of my research is to optimize small molecule inhibitor SE-1. Despite the fact that SE-1 has been shown to block DNA binding by many AraC family proteins, SE-1 showed some nonspecific inhibition at higher concentrations against non-AraC family proteins CRP and LacI. In addition, the concentration of SE-1 required to inhibit AraC family activators would need to be decreased for SE-1 to be useful as a novel antimicrobial agent. In collaboration with Dr. Jeff Aubé, I assayed both commercially purchased analogs of SE-1 and analogs synthesized by the Aubé group in the whole cell based *in vivo* inhibitor assay. Unfortunately, no analogs were found increased VirF inhibition potency relative to SE-1. However, in the process of synthesizing analogs, we found that SE-1 and its analogues converted to corresponding quaternary salts in aqueous solution, and the quinolinium salt was responsible for the observed inhibition by SE-1. In addition, by testing the inhibition of SE-1 and its analogues in EMSA with purified RhaR protein, we found that some uncharged analogues show more inhibition *in vitro* than *in vivo*, suggesting that the uptake of the analogues into bacterial cells may be a reason why some analogues did not show inhibition in the whole cell-based assay.

Purification of RhaR and Identification of RhaR annotation error. To better understand the properties of RhaR, and the mechanism of transcription activation by RhaR, the fourth goal of my research was to purify RhaR protein to perform biochemical assays. In initial purifications of RhaR, I found that there were two different sized RhaR proteins being purified, and DNA-binding assays suggested that the smaller sized RhaR was the active form. Protein peptide mapping and N-terminal sequencing indicated that the smaller sized RhaR was missing the 30 amino acids relative to the larger sized RhaR. Further examination of the sequence of the *rhaR* open reading frame showed that immediately upstream of the sequence encoding the N-terminal protein sequence was a GTG codon, which can also be used as a start codon in *E. coli*. To test the hypothesis that the GTG codon served as the true *rhaR* start codon, I constructed a new plasmid to test the putative GTG start codon and its associated Shine-Dalgarno (SD) sequence, and tested the ability of the encoded RhaR to activate transcription in a reporter gene assay. The results support the hypothesis that wild-type RhaR translation initiates at the GTG codon, not the upstream ATG codon, and that the previously annotated RhaR protein sequence had 30 extra amino acids at its N-terminal end relative to the native 282 residue RhaR protein sequence.

High Resolution Structures of the Regulatory Domain of RhaR. To provide more structural information about RhaR protein, the fifth goal of my research is to crystalize RhaR-NTD and solve the structure of RhaR-NTD. After crystallization screening and condition optimization, I was able to obtain RhaR-NTD crystals, and the structures of RhaR-NTD were determined. The rhamnose-bound RhaR-NTD structure showed that the protein forms an antiparallel dimer, and shared a fold that was similar to the AraC-NTD, binding its respective sugar L-rhamnose within a β-barrel. In addition, a Ni^{2+} ion, which has not been seen in other AraC family protein structures, was present in the sugar-binding, cupin superfamily, motif of RhaR. A rhamnose-free structure

was also solved, and in this structure, a loop region that is involved in rhamnose binding was completely disordered. A second loop region also has minor structural changes. Each of the two regions with rhamnose-dependent structural changes is predicted to be at the interface between the RhaR NTD and DNA-binding domain, suggesting their potential involvement in rhamnose allosteric signaling. No differences were observed in the RhaR N-terminal arm region in the fully and partially rhamnose-occupied structures, suggesting that RhaR rhamnose-dependent allosteric signaling shares some features with the 'light switch' model of AraC, but differs in other features.

DNA binding and allosteric signaling by RhaR. The last goal of my research was to further characterize the mechanism by which RhaR responds to L-rhamnose. I have measured the effects of rhamnose on the equilibrium binding affinity of RhaR to the RhaR protomer (half) and dimer (full) DNA binding sites, and determined the rhamnose dependence of DNA bending by RhaR. The equilibrium binding affinity of the isolated RhaR-DBD for the half DNA binding site was also measured. RhaR binding affinity to the half site DNA was independent of rhamnose and 300-fold tighter than for the DBD alone. In contrast, rhamnose increased the strength and angle of bending of RhaR bound to full site DNA. The findings support the hypothesis that RhaR NTD increased the DNA binding affinity of each RhaR protomer, independent of rhamnose; and that rhamnose signalling primarily increased the cooperativity of binding to the full site DNA by the two DBDs in a RhaR dimer. These findings enable development of a hypothesis regarding allosteric rhamnose signalling in RhaR.

Chapter 2

MATERIALS AND METHODS

Growth conditions. *Escherichia coli* strains were grown in tryptone-yeast extract (TY) broth (0.8% Difco tryptone, 0.5% Difco yeast extract, and 0.5% NaCl, [pH 7.0]) unless otherwise indicated. Difco nutrient agar (1.5% agar) (BD, Cockeysville, MD) was used routinely to grow *E. coli* strains on solid medium. To test for *lacZ* expression by RhaS and RhaR on solid media, Difco nutrient agar was supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, 40 μg/ml) and X-gal with L-rhamnose (0.2%), respectively. All cultures were grown at 37°C, and liquid cultures were grown with shaking.

General Methods. Primers used in this study were synthesized by Eurofins MWG Operon (Huntsville, AL). PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). Amplified DNA fragments were cleaned up using either E.Z.N.A. Cycle Pure (Omega Bio-tek, Norcross, GA) or IBI Gel/PCR products extraction kits (IBI Scientific, Peosta, IA). Standard methods were followed for restriction endonuclease digestion, ligation, and transformation. Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Plasmid isolation from bacteria was performed using the IBI High-Speed Plasmid Mini kit (MIDSCI, St. Louis, MO). DNA sequencing was performed to obtain sequences of for both strands of the entire cloned region for all cloned and mutagenized DNA fragments at the Northwestern University Genomics Core (Chicago, IL).

Bacterial strains, plasmids and oligos. Bacterial strains, plasmids and oligos used in this study are shown in Tables 1, 2, 3 respectively. All strains used in β-galactosidase assays were derived from ECL116 [85].

Table 1. Bacterial strains used in this study.

Strain	Genotype	Source or Reference
SME1088	ECL 116 *ΔS recA*::*cat* λΦ 772 *rha*B-lacZ Δ84	Laboratory collection
SME1222	ECL 116 *ΔrhaS*/ λBS 742 (*rhaB*-lacZ Δ110)	Laboratory collection
SME1773	SME1222 + pALTER-1-mut *rhaS* F248A	This study
SME1774	SME1222 + pALTER-1-mut *rhaS* H253A	This study
SME2525	ECL116 λΦ(*rhaS*-lacZ) Δ128 Δ*rhaSR*::*kan recA*::*cat*	Laboratory collection
SME3160	ECL116 λΦ(*rha*S-*lacZ*) Δ85 *ΔrhaSR*::*kan recA*::*cat*	Laboratory collection
SME3358	SME1085(Δ*rhaS recA*::*cat*) / λHTS1-*lacZ*	[81]
SME3632	SME3600 *recA*::*cat*, final genotype = λΦ(*rhaB-lacZ*)Δ84 Δ(*rhaSR*)::*kan malP*::*lacI*^{q} *zhc-511*::Tn*10 recA*::*cat*	[81]
SME3359	SME3358 + pHG165 *lacI*	This study
SME3634	SME3632 + pHG165 *rhaS*	This study
SME4259	ECL116 λMAD102Φ[(*virB-lacZ*)] *malP*::*lacI*^{q} *zhc-511*::Tn*10 recA*::*cat*	[81]
SME4384	SME1088+ pHG165 *rhaS* V188A (Amp^{R})	This study
SME4385	SME1088+ pHG165 *rhaS* W190A (Amp^{R})	This study
SME4386	SME1088+ pHG165 *rhaS* D191A (Amp^{R})	This study
SME4387	SME1088+ pHG165 *rhaS* D191R (Amp^{R})	This study
SME4388	SME1088+ pHG165 *rhaS* L220A (Amp^{R})	This study
SME4389	SME1088+ pHG165 *rhaS* L220R (Amp^{R})	This study
SME4390	SME1088+ pHG165 *rhaS* H253A/ (Amp^{R})	This study
SME4395	SME1088+ pHG165 *rhaS* V188R (Amp^{R})	This study
SME4396	SME1088+ pHG165 *rhaS* N221A (Amp^{R})	This study

SME4397	SME1088+ pHG165 *rhaS* N221R (Amp^R)	This study
SME4398	SME1088+ pHG165 *rhaS* S249R (Amp^R)	This study
SME4399	SME1088+ pHG165 *rhaS* L257A (Amp^R)	This study
SME4400	SME1088+ pHG165 *rhaS* L257R (Amp^R)	This study
SME4414	SME1088+ pHG165 *rhaS* R224A (Amp^R)	This study
SME4415	SME1088+ pHG165 *rhaS* F248A (Amp^R)	This study
SME3564	SME3160 + pHG165 *rhaR*	This study
SME4601	BL21 (DE3) + pSE308 (pET24b *rhaR-his*$_6$) (Kan^R)	This study
SME2483	BL21 (DE3) + pSE326 (pET24b +30 *rhaR-his*$_6$) (Kan^R)	This study
SME4608	SME3160 + pHG165 *rha*R (Amp^R)	This study
SME4612	SME3160 + pSE312(pHG165 *rhaR*)	This study
SME4614	SME3160 + pSE313 (pHG165 *rhaR*)	This study
SME4666	BL21(DE3) +pSE226 (pET24b *rhaR*-DBD-*his*$_6$) (Kan^R)	This study
SME4668	BL21(DE3) +pSE327 (pET24b *rhaR*-NTD-*his*$_6$) (Kan^R)	This study
SME4673	B834(DE3) + pSE327 (pET24b *rhaR*-NTD-*his*$_6$) (Kan^R)	This study
SME3565	SME2525 + pHG165 *rhaR* (Amp^R)	This study
SME4677	SME2525 + pHG165 *rhaR* V20A (Amp^R)	This study
SME4678	SME2525 + pHG165 *rhaR* Q28A (Amp^R)	This study
SME4679	SME2525 + pHG165 *rhaR* H34A (Amp^R)	This study
SME4680	SME2525 + pHG165 *rhaR* H74A (Amp^R)	This study
SME4681	SME2525 + pHG165 *rhaR* Y76A (Amp^R)	This study
SME4687	SME2525 + pHG165 *rhaR* R25A (Amp^R)	This study
SME4688	SEM2525 + pHG165 *rhaR* F31A (Amp^R)	This study

SME4689	SEM2525 + pHG165 *rhaR* N86A (Amp^R)	This study
SME4382	SME4259 + pHG165 *virF*	This study
SME4706	SME3632 + pHG165 *rhaS* V188A (Amp^R)	This study
SME4707	SME3632 + pHG165 *rhaS* D191A (Amp^R)	This study
SME4708	SME3632+ pHG165 *rhaS* D191R (Amp^R)	This study
SME4709	SME3632+ pHG165 *rhaS* L220A (Amp^R)	This study
SME4710	SME3632+ pHG165 *rhaS* L220R (Amp^R)	This study
SME4711	SME3632+ pHG165 *rhaS* N221A (Amp^R)	This study
SME4712	SME3632+ pHG165 *rhaS* N221R (Amp^R)	This study
SME4713	SME3632+ pHG165 *rhaS* S249R (Amp^R)	This study
SME4714	SME3632+ pHG165 *rhaS* H253A (Amp^R)	This study
SME4715	SME3632+ pHG165 *rhaS* L257A (Amp^R)	This study
SME4716	SME3632+ pHG165 *rhaS* L257R (Amp^R)	This study
SME 2923	BL21 (DE3) + pSE230 (pET15b *His_6-rhaS CTD*)	This study
SME4618	BL21 (DE3) + pSE328 (pET24b *rhaS-his_6*) (Kan^R)	This study
SME4631	BL21 (DE3) + pSE329 (pET24b *rhaS*L201R-*his_6*) (Kan^R)	This study
SME4037	SME1088 + pHG165 *rhaS* L208M	This study
SME4038	SME1088 + pHG165 *rhaS* L208I	This study
SME4039	SME1088 + pHG165 *rhaS* Q210R	This study
SME4040	SME1088 + pHG165 *rhaS* Q210S	This study
SME4041	SME1088 + pHG165 *rhaS* Q210G	This study
SME4042	SME1088 + pHG165 *rhaS* Q210N	This study
SME4043	SME1088 + pHG165 *rhaS* Q210K	This study

SME4044	SME1088 + pHG165 *rhaS* Q210G	This study
SME4045	SME1088 + pHG165 *rhaS* Q210V	This study
SME4046	SME1088 + pHG165 *rhaS* Q210I	This study
SME4047	SME1088 + pHG165 *rhaS* T212C	This study
SME4048	SME1088 + pHG165 *rhaS* T212M	This study
SME4049	SME1088 + pHG165 *rhaS* T212A	This study
SME4050	SME1088 + pHG165 *rhaS* L214V	This study
SME4051	SME1088 + pHG165 *rhaS* L214I	This study
SME4052	SME1088 + pHG165 *rhaS* L214M	This study
SME4053	SME1088 + pHG165 *rhaS* L214G	This study
SME4054	SME1088 + pHG165 *rhaS* R218V	This study
SME4055	SME1088 + pHG165 *rhaS* R218L	This study
SME4056	SME1088 + pHG165 *rhaS* R218G	This study
SME4057	SME1088 + pHG165 *rhaS* Y219F	This study
SME4058	SME1088 + pHG165 *rhaS* Y219L	This study
SME4059	SME1088 + pHG165 *rhaS* Y219F	This study
SME4060	SME1088 + pHG165 *rhaS* M226L	This study
SME4061	SME1088 + pHG165 *rhaS* M226E	This study
SME4062	SME1088 + pHG165 *rhaS* M226R	This study
SME4063	SME1088 + pHG165 *rhaS* M226V	This study
SME4064	SME1088 + pHG165 *rhaS* M226G	This study
SME4065	SME1088 + pHG165 *rhaS* M226D	This study
SME4066	SME1088 + pHG165 *rhaS* K227Y	This study

SME4067	SME1088 + pHG165 *rhaS* K227F	This study
SME4068	SME1088 + pHG165 *rhaS* K227D	This study
SME4069	SME1088 + pHG165 *rhaS* K227L	This study
SME4070	SME1088 + pHG165 *rhaS* K227S	This study
SME4071	SME1088 + pHG165 *rhaS* K227S	This study
SME4072	SME1088 + pHG165 *rhaS* R229I	This study
SME4073	SME1088 + pHG165 *rhaS* R229G	This study
SME4074	SME1088 + pHG165 *rhaS* R229D	This study
SME4075	SME1088 + pHG165 *rhaS* R229T	This study
SME4076	SME1088 + pHG165 *rhaS* R229P	This study
SME4077	SME1088 + pHG165 *rhaS* H230E	This study
SME4078	SME1088 + pHG165 *rhaS* H230V	This study
SME4079	SME1088 + pHG165 *rhaS* H230I	This study
SME4080	SME1088 + pHG165 *rhaS* H230Q	This study
SME4081	SME1088 + pHG165 *rhaS* H230L	This study
SME4082	SME1088 + pHG165 *rhaS* H230G	This study
SME4083	SME1088 + pHG165 *rhaS* H230R	This study
SME4084	SME1088 + pHG165 *rhaS* L232M	This study
SME4085	SME1088 + pHG165 *rhaS* L232V	This study
SME4086	SME1088 + pHG165 *rhaS* L232I	This study
SME4087	SME1088 + pHG165 *rhaS* R233L	This study
SME4088	SME1088 + pHG165 *rhaS* R233P	This study
SME4089	SME1088 + pHG165 *rhaS* R233A	This study

SME4090	SME1088 + pHG165 *rhaS* R233W	This study
SME4091	SME1088 + pHG165 *rhaS* R233D	This study
SME4092	SME1088 + pHG165 *rhaS* R233G	This study
SME4093	SME1088 + pHG165 *rhaS* H234T	This study
SME4094	SME1088 + pHG165 *rhaS* H234I	This study
SME4095	SME1088 + pHG165 *rhaS* H234K	This study
SME4096	SME1088 + pHG165 *rhaS* H234G	This study
SME4097	SME1088 + pHG165 *rhaS* H234Y	This study
SME4098	SME1088 + pHG165 *rhaS* H234V	This study
SME4099	SME1088 + pHG165 *rhaS* H234D	This study
SME4100	SME1088 + pHG165 *rhaS* H234E	This study
SME4101	SME1088 + pHG165 *rhaS* E261N	This study
SME4102	SME1088 + pHG165 *rhaS* E261G	This study
SME4103	SME1088 + pHG165 *rhaS* E261S	This study
SME4104	SME1088 + pHG165 *rhaS* E261M	This study
SME4105	SME1088 + pHG165 *rhaS* E261F	This study
SME4106	SME1088 + pHG165 *rhaS* F262Y	This study
SME4107	SME1088 + pHG165 *rhaS* F262L	This study
SME4108	SME1088 + pHG165 *rhaS* N263Y	This study
SME4109	SME1088 + pHG165 *rhaS* N263E	This study
SME4110	SME1088 + pHG165 *rhaS* N263R	This study
SME4111	SME1088 + pHG165 *rhaS* N263L	This study
SME4112	SME1088 + pHG165 *rhaS* N263W	This study

SME4113	SME1088 + pHG165 *rhaS* W264H	This study
SME4114	SME1088 + pHG165 *rhaS* W264M	This study
SME4115	SME1088 + pHG165 *rhaS* W264E	This study

Table 2. Plasmids used in this study.

Plasmid	Relevant characteristics	Source or Reference
pHG165	*lacZα rop* Amp^R (ColE1 origin from pBR322)	[86]
pET15b	*lacI* Amp^R (ColE1 origin from pBR322)	Novogen
pET24b	*lacI* Kan^R (ColE1 origin from pBR322)	Novogen
pTYB11	*lacI rop* Amp^R (pMB1 origin)	New England BioLabs
pTXB1	*lacI rop* Amp^R (pMB1 origin)	New England BioLabs
pSE325	pTYB11 +30*rhaR*	This study
pSE326	pET24b +30*rhaR-his*$_6$	This study
pSE230	pET15b *his*$_6$*-rhaS* 163-278 (DBD)	This study
pSE308	pET24b *rhaR-his*$_6$	This study
pSE327	pET24b *rhaR* 1-172 (NTD) -*his*$_6$	This study
pSE228	pET24b *rhaR* 166-282 (DBD) -*his*$_6$	This study
pRT	pTXB1 *toxT*	R. Taylor
pSE302	pHG165 +30*rhaR*	This study

pSE312	pHG165 *rhaR* (GTG start codon)	This study
pSE313	pHG165 +30*rhaR-his*$_6$ (extra 30-residue at N-terminus, extra nucleotide after ATG, out of frame expression)	This study
pHG165 *lacI*	*lacI* expressed from pHG165 *lac* promoter, Amp^R	[81]
pSE328	pET24b *rhaS-his*$_6$	This study
pSE329	pET24b *rhaS*L201R-*his*$_6$	This study

Table 3. Oligonucleotides used in this study.

Oligo No.	Oligonucleotide sequence 5'-3'	Use
2082	GTGAAGCTTTTAATCTTTCTGCGAATTGAG	Clone *rhaR* into pHG165 (Downstream)
2098	GATGAATTCGCCGTGTTGACGACATCAGGA	Clone *rhaS* into pHG165 (Upstream)
2344	GATGATCTCGAGGCTACCACGCGGAACCAGGGATGTTG GCGGCAACG	Clone *rhaR* NTD into pET24b (Downstream)
2345	GCGCCGCGCATATGAGTGATTCGTTGCCGCCA	Clone *rhaR* DBD into pET24b (Upstream)
2349	GCGCCGCGCATATGCAGGAGAACCTGGAAAACAGC	Clone *rhaS* and *rhaS*L201R into pET24b (Upstream)
2349	GCGCCGCGCATATGCAGGAGAACCTGGAAAACAGC	Clone *rhaS* DBD into pET15b or pET24b (Upstream)
2352	GATGATCTCGAGGCTACCACGCGGAACCAGATCTTTCT GCGAATTGAGATGAC	Clone *rhaR* or *rhaR* DBD into pET24b (Downstream)
2353	GATGATCTCGAGGCTACCACGCGGAACCAGTTGCAGAA AGCCATCCCG	Clone *rhaS* and *rhaS*L201R into pET24b (Downstream)
2714	TGAGTAAAGCTTTTATTGCAGAAAGCCATCCCG	Clone *rhaS* into pHG165 (Downstream)
2967	CGGGATCCTTATTGCAGAAAGCCATCCC	Clone *rhaS* or *rhaS* DBD into pET15b (Downstream)
3355	GGGCTCTTCGGCAAAATGGTCCTCCAGCCAGGCCAGAA GCAAGTTG	Construct *rhaS* V188A, V188R, W190A, D191A, D191R (Downstream)
3356	GGGCTCTTCTTGCCGATGAGGCGAATTGGGATGCCGTG GCGGAT	Construct *rhaS* V188A (Upstream)
3357	GGGCTCTTCTTGCCGATGAGGTGAATGCGGATGCCGTG GCGGAT	Construct *rhaS* W190A (Upstream)
3358	GGGCTCTTCTTGCCGATGAGGTGAATTGGGCGGCCGTG GCGGAT	Construct *rhaS* D191A (Upstream)
3359	GGGCTCTTCTTGCCGATGAGCGTAATTGGGATGCCGTG GCGGAT	Construct *rhaS* V188R (Upstream)

3360	GGGCTCTTCTTGCCGATGAGGTGAATTGGCGTGCCGTGGCGGAT	Construct *rhaS* D191R (Upstream)
3361	GGGCTCTTCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCG	Construct *rhaS* L220A, L220R, N221A, N221R, R224A (Downstream)
3362	GGGCTCTTCCCTCAGCGATACGCGAACCGCCTGCGACTGATGAAAG	Construct *rhaS* L220A (Upstream)
3363	GGGCTCTTCCCTCAGCGATACCGTAACCGCCTGCGACTGATGAAAG	Construct *rhaS* L220R (Upstream)
3364	GGGCTCTTCCCTCAGCGATACCTGGCGCGCCTGCGACTGATGAAAG	Construct *rhaS* N221A (Upstream)
3365	GGGCTCTTCCCTCAGCGATACCTGCGTCGCCTGCGACTGATGAAAG	Construct *rhaS* N221R (Upstream)
3366	GGGCTCTTCCCTCAGCGATACCTGAACCGCCTGGCGCTGATGAAAG	Construct *rhaS* R224A (Upstream)
3367	GGGCTCTTCTCCACAGCGATAGGCGATGTCAGTAAC	Construct *rhaS* S249R, L257A, L257R (Downstream)
3368	GGGCTCTTCGTGGATTCCGTGACAGTAACCACTTTTCGACGCTTTTTCGCCGAGAGTTTA	Construct *rhaS* S249A (Upstream)
3369	GGGCTCTTCGTGGATTCAGCGACAGTAACCACTTTTCGACGGCGTTTCGCCGAGAGTTTA	Construct *rhaS* L257A (Upstream)
3370	GGGCTCTTCGTGGATTCAGCGACAGTAACCACTTTTCGACGCGUTTTCGCCGAGAGTTTA	Construct *rhaS* L257R (Upstream)
3415	GCGCCGCGCATATGGCGCATCAGTTAAAACTTCTC	Clone *rhaR* into pET24b, pET15b, and pTYB1 (Upstream)
3416	CTGGCGGCTCTTCCAACGCGCATCAGTTAAAACTTCTC	Clone *rhaR* into pTYB11 (Upstream)
3417	GATGAATTCGGTCACCGCGTGATATTCGCCAGGGACGGGATGGCGCATCAGTTAAAACTTC	Clone *rhaR* into pHG165 (Upstream)
3430	GATGAATTCCAACTTCGTTCTCTGGCCGAGGTAGCCACGGTGGCGC	Clone *rhaR* into pHG165 (to test a possible downstream SD) (Upstream)
3431	GATGAATTCGGTCACCGCGTGATATTCGCCAGGGACGGGATGAGCTTTC	Clone *rhaR* into pHG165 (to introduce a frame shift) (Upstream)

3056	[IRD700]CGCTGTATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTAAGAAC	IRD700 labeled (MWG) top strand with full RhaR DNA binding site for EMSA
3057	[IRD700]GTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCG	IRD700 labeled (MWG) bottom strand with full RhaR DNA binding site for EMSA
3457	[aminoC6+DY682]TTTCCGTCGAAAATTTAAGGTAAGA	IR682 labeled (MWG) top strand with RhaR half site (*rhaI*$_4$) for EMSA
3459	TCTTACCTTAAATTTTCGACGGAAA	Bottom strand with RhaR half site (*rhaI*$_4$) for EMSA
3058	[IRD700]ACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAA	IRD700 labeled (MWG) top strand with full RhaS DNA binding site for EMSA
3288	TTCGCGACCTTCTCGTTACTGACAGGAAAATGGGCCATTGGCAACCAGGGAAAGATGAACGT	Bottom stand with full RhaS DNA binding site for EMSA
3460	[aminoC6+DY682]GTTCATCTTTCCCTGGTTGCCAATG	IR682 labeled (MWG) top strand with *rhaI*$_1$ DNA binding site for EMSA
3461	CATTGGCAACCAGGGAAAGATGAAC	Bottom strand with *rhaI*$_1$ DNA binding site for EMSA
3401	GTGTTATTAAAAAAATAAAAAAACACAGCAAAAAATGACA	Top strand of ToxT binding site on *tcp* promoter for EMSA. From [38]
3402	TGTCATTTTTTGCTGTGTTTTTTTATTTTTTTAATAACACCCAGACCAGGGCAC	Bottom strand of ToxT binding site on *tcp* promoter for EMSA. From [38]
3371	[aminoC6+DY682]GTGCCCTGGTCTGG	IR700 labeled oligo for LUEGO IR labeling of DNA for EMSAs
3481	GTCCTCTTCGCAGGCAGCGGCTGTGGCTGACCGTTATCCG	Construct *rhaR* V20A (Upstream)
3482	GTCCTCTTCCCTGCTGGTCGCTGGCAAAAAAATCATCTTT	Construct *rhaR* V20A (Downstream)
3483	GTCCTCTTCATCCGGCGGATGTCTTTGCTGAACATACACA	Construct *rhaR* Q28A (Upstream)

3484	GTCCTCTTCATCCGCAAGATGTCTTTGCTGAAGCGACA CATGATTTTTGTGAGCTG	Construct *rhaR* H34A (Upstream)
3485	GTCCTCTTCCGGATAACGGTCAGCCACAGCGACTGCCT GCTGG	Construct *rhaR* Q28A, F31A, H34A (Downstream)
3486	GTCCTCTTCCGATAAAGCGTCCTACGCTTCCGTTAACG ATCTGGTTTTG	Construct *rhaR* H74A (Upstream)
3487	GTCCTCTTCCGATAAACACTCCGCGGCTTCCGTTAACG ATCTGGTTTTG	Construct *rhaR* Y76A (Upstream)
3488	GTCCTCTTCTATCGTCAGCATGAATGTAAAAGAGATCG CCACGGGTAATGC	Construct *rhaR* H74A and Y76A (Downstream)
3489	GTCCTCTTCCTGACGCGTATCCGCAAGATGTCTTTGCT GA	Construct *rhaR* R25A (Upstream)
3490	GTCCTCTTCGTCAGCCACAGCGACTGCCTGCTGGTCGC TG	Construct *rhaR* R25A (Downstream)
3491	GTCCTCTTCATCCGCAAGATGTCGCGGCTGAACATACA CA	Construct *rhaR* F31A (Upstream)
3492	GTCCTCTTCTGCAGGCGATTATTTATTGCCCGGAGCGT CT	Construct *rhaR* N86A (Upstream)
3493	GTCCTCTTCCTGCAAAACCAGATCGTTAACGGAAGCGT AG	Construct *rhaR* N86A (Downstream)
3433	ATCACAGTGAGACTGACAGT	Downstream primer to amplify DNA for bending experiment
3101	GTGCTCTTCACATCGGCAG**NNN**AAGCAGCAAACG	Construct *rhaS* L208X (Upstream)
3102	GTGCTCTTCGATGTAGCGTACGCAGTGAAAG	Construct *rhaS* L208X (Downstream)
3103	GTGCTCTTCTAAGCAG**NNN**ACGGGACTGACG	Construct *rhaS* Q211X (Upstream)
3104	GTGCTCTTCTAAGCAGCAA**NNN**GGACTGACGCCTCAG	Construct *rhaS* T212X (Upstream)
3105	GTGCTCTTCTAAGCAGCAAACGGGA**NNN**ACGCCTCAGC GATAC	Construct *rhaS* L214X (Upstream)

3106	GTGCTCTTCACTTAAGCTGCCGATGTAGCG	Construct *rhaS* Q211X, T212X, L214X (Downstream)
3107	GTGCTCTTCGCCTCAG**NNN**TACCTGAACCGCCTG	Construct *rhaS* R218X (Upstream)
3108	GTGCTCTTCGCCTCAGCGA**NNN**CTGAACCGCCTGCG	Construct *rhaS* Y219X (Upstream)
3109	GTGCTCTTCGAGGCGTCAGTCCCGTTTG	Construct *rhaS* R218X, Y219X (Downstream)
3190	AGTCTCTTCTCGCAGGCGGTTCAGGTATCGC	Construct *rhaS* M226X (Downstream)
3191	AGTCTCTTCTGCGACTG**NNN**AAAGCCCGACATCTGCTACGCC	Construct *rhaS* M226X (Upstream)
3192	AGTCTCTTCGCAGCTTAAG**NNN**CAAACGGGACTGACG	Construct *rhaS* Q210X (Upstream)
3193	AGTCTCTTCGCTGCCGATGTAGCGTACGCAG	Construct *rhaS* Q210X (Downstream)
3194	AGTCTCTTCTCGCCGA**NNN**TTTAACTGGTCACCGCGTGATATTC	Construct *rhaS* E261X (Upstream)
3195	AGTCTCTTCTCGCCGAGAG**NNN**AACTGGTCACCGCGTGATATTC	Construct *rhaS* F262X (Upstream)
3196	AGTCTCTTCTCGCCGAGAGTTT**NNN**TGGTCACCGCGTGATATTC	Construct *rhaS* N263X (Upstream)
3197	AGTCTCTTCTCGCCGAGAGTTTAAC**NNN**TCACCGCGTGATATTC	Construct *rhaS* W264X (Upstream)
3198	AGTCTCTTCCGCGAAAAAGCGTCGAAAAGTGGTTAC	Construct *rhaS* E261X,F262X,N263X,W264X (Upstream)
2946	AGTCTCTTCACAGTCGCAGGCGGTTCAGG	Construct *rhaS* K227X, R229X, H230X (Downstream)

2947	AGTCTCTTCACTGATG**NNN**GCCCGACATCTGC	Construct *rhaS* K227X (Upstream)
2948	AGTCTCTTCACTGATGAAAGCCCGA**NNN**CTGCTACGCC ACAG	Construct *rhaS* H230X (Upstream)
2949	AGTCTCTTCATAGCAGATGTCGGGCTTTC	Construct *rhaS* R233X, H234X (Downstream)
2950	AGTCTCTTCACTA**NNN**CACAGCGAGGCCAG	Construct *rhaS* R233X (Upstream)
2951	AGTCTCTTCACTACGC**NNN**AGCGAGGCCAGCG	Construct *rhaS* H234X (Upstream)
2850	AGTTCACTCTTCGGGCTTTCATCAGTCGCAGG	Construct *rhaS* L232X (Downstream)
2851	AGCTCACTCTTCAGCCCGACATCTG**NNN**CGCCACAGCG	Construct *rhaS* L232X (Upstream)
3100	AGTCTCTTCACTGATGAAAGCC**NNN**CATCTGCTACGCC ACAG	Construct *rhaS* R229X (Upstream)

Underlined sequences are primer-template mismatches that include restriction endonuclease cleavage sites and random nucleotides.

For oligonucleotides with **NNN**, **N**= A, G, C, T

DNAs for RhaR DNA bending experiments. To measure the angle of RhaR DNA bending, we designed five fragments of DNA with RhaR full binding site at various positons. The DNA fragments were subsequently synthesized and cloned to plasmid pCR2.1by Eurofins MWG Operon (Huntsville, AL). PCR was used to amplify and label each DNA fragment. The set of oligos used was 3371 -3433, and the DNA sequences of each fragment are listed below. The RhaR full binding site was underlined.

Fragment 1:

GTGCCCTGGTCTGGCACTGCATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGG
TACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAACATCGATCAGCTTATCCAC
CATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG
CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGT

Fragment 2:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTATCGATCAGCTTATCCAC
CATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG
CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGT

Fragment 3:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTATCTTGAAAAATCGACGT
TTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG
CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGT

Fragment 4:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACAC
ATATGGAATCACTTGTCAGCTGTCTAGACACAGATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGT
CGAAAATTTAAGGTCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGT

Fragment 5:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACAC
ATATGGAATCACTTGTCAGCTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCG
CGAGTCGACTGTGAATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTACTGT
CAGTCTCACTGT

DNAs for RhaS DNA bending experiments. To measure the angle of RhaS DNA bending, we designed five fragments of DNA with RhaS full binding site at various positons. The DNA fragments were subsequently synthesized and cloned to plasmid pCR2.1by Eurofins MWG Operon (Huntsville, AL). PCR was used to amplify and label each DNA fragment. The set of oligos used was 3371 -3433, and the DNA sequences of each fragment are listed below. The RhaS full binding site was underlined.

Fragment 1:

GTGCCCTGGTCTGGCACTGCATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGG
TACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAACATCGATCAGCTTATCCAC
CATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG

CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGTGAT

Fragment 2:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTATCGATCAGCTTATCCAC
CATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG
CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGTGAT

Fragment 3:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTATCTTGAAAAATCGACGT
TTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTCTCGAGGGATCCTGACACATATGGAATCACTTGTCAG
CTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGTGAT

Fragment 4:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACAC
ATATGGAATCACTTGTCAGCTGTCTAGACACAGATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGT
CGAAAATTTAAGGTCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCGCGAGTCGACTGTGAACTGT
CAGTCTCACTGTGAT

Fragment 5:

GTGCCCTGGTCTGGCACTGCACGCGTGGTACCACGATAGATCTCGAGTCGTCAGGCTAGCCCATGGACAA
CATCGATCAGCTTATCCACCATGGATCGAGTCAGTACTAGTCGCTGGATGTTCTCGAGGGATCCTGACAC
ATATGGAATCACTTGTCAGCTGTCTAGACACAGCCCGGGACTCAGCTCGTAGGCCTACGCACTCAGTTCG
CGAGTCGACTGTGAATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTACTGT
CAGTCTCACTGTGAT

Construction of RhaS variants to identify SE-1 binding site. Wild-type *rhaS* and all variants were cloned into and expressed from the plasmid pHG165 [86] using the restriction enzymes EcoRI and HindIII. Wild type *rhaS, rhaS* F248A*, and rhaS* H253A were PCR amplified using oligos 2098 and 2714. Site-specific mutagenesis of *rhaS* was performed using PCR to make oligonucleotide-directed mutations at each of the desired positions. Wild type *rhaS* was used as template for all PCR amplifications. The specific oligo sets used to create each set of mutants are described in Table 4. Plasmids carrying mutations were then transformed into strain SME1088 for initial evaluation of the transcription activity, and then were transformed into SME3632 for testing SE-1 analogues *in vivo* dose-dependent assay.

Table 4. Oligonucleotides to Amplify Fragments to Generate RhaS Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
V188A	2098-3355	3356-2714
V188R	2098-3355	3359-2714
W190A	2098-3355	3357-2714
D191A	2098-3355	3358-2714
D191R	2098-3355	3360-2714
L220A	2098-3361	3362-2714
L220R	2098-3361	3363-2714
N221A	2098-3361	3364-2714
N221R	2098-3361	3365-2714
R224A	2098-3361	3366-2714
S249R	2098-3367	3368-2714
L25RA	2098-3367	3369-2714
L257R	2098-3367	3370-2714

Construction of RhaR L-rhamnose binding variants. Variants were cloned using standard techniques as described above. The specific oligo sets used to create each variant are described in Table 5. Plasmids carrying mutations were transformed into strain SME2525 for β-galactosidase assay.

Table 5. Oligonucleotides to Amplify Fragments to Generate RhaR Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
V20A	3430-3482	3481-2082
Q28A	3430-3485	3483-2082
F31A	3430-3485	3491-2082
H34A	3430-3485	3484-2082
H74A	3430-3488	3486-2082
Y76A	3430-3488	3487-2082
N86A	3430-3493	3492-2082

Construction of RhaS variants to elucidate RhaS rhamnose allosteric signaling. Wild-type *rhaS* and all variants were cloned into and expressed from the plasmid pHG165 [86] using the restriction enzymes EcoRI and HindIII. Random mutations were introduced by amplifying *rhaS* using a pair of oligonucleotides with the downstream oligonucleotide carrying a randomized codon (NNN) at the desired position with wild type template. Wild type *rhaS* was used as template for all PCR amplifications. The specific oligo sets used to create each set of mutants are described in Table 6. Plasmids carrying mutations were then transformed into strain SME1088 for evaluation of the transcription activity using β-galactosidase assays.

Table 6. Oligonucleotides to Amplify Fragments to Generate RhaS Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
L208X	2098-3102	3101-2714

Q210X	2098-3193	3192-2714
Q211X	2098-3106	3103-2714
T212X	2098-3106	3104-2714
L214X	2098-3106	3105-2714
R218X	2098-3109	3107-2714
Y219X	2098-3109	3108-2714
M226X	2098-3190	3191-2714
K227X	2098-2946	2947-2714
R229X	2098-2946	3100-2714
H230X	2098-2946	2948-2714
L232X	2098-2850	2851-2714
R233X	2098-2949	2950-2714
H234X	2098-2949	2951-2714
E261X	2098-3198	3194-2714
F262X	2098-3198	3195-2714
N263X	2098-3198	3196-2714
W264X	2098-3198	3197-2714

Construction of RhaR expression plasmids to test upstream start codons and Shine-Dalgarno (SD) sequences for *in vivo* experiment. In order to test the native RhaR start codon GTG, *rhaR* (282 residues) with associated SD sequence was cloned into pHG165 to make plasmid pSE312 using oligos 3430 and 2082. To introduce a frame shift at the upstream ATG start codon of *rhaR* (312 residues), oligos 3031 and 2082 were used to create plasmid pSE313. The cloning method consisted of PCR amplification of the gene encoding different size of RhaR and their corresponding SD. The PCR fragments and plamid pHG165 were both digested with restriction endonucleases EcoRI and HindIII, and then ligated with T4 ligase.

Construction of plasmids for overexpression of RhaR-His_6, RhaR-NTD-His_6, and RhaR-DBD-His_6 proteins. Plasmids were constructed for overexpression of RhaR-His_6, RhaR-NTD (residue 1-172)-His_6 RhaR-DBD (residue 166-282)-His_6. The inserts encoding RhaR were generated by PCR with oligos 2352 and 3415. The RhaR-NTD insert was generated by PCR with oligos 3451 and 2344. The RhaR-DBD fragment was amplified by PCR with oligos 2345

and 2352. For all the inserts, a thrombin cleavage site was primer encoded at the C-terminal end of the protein. All the inserts were digested with restriction endonucleases NdeI and XhoI, and then ligated with similarly digested plasmid pET24b to make pSE308 (encodes RhaR-His_6), pSE327 (encodes RhaR-NTD-His_6), and pSE226 (encodes RhaR-DBD-His_6).

Construction of plasmid for overexpression of RhaS-His_6, and $RhaS^{L201R}$-His6. Plasmids were constructed for overexpression of RhaS-His_6 and $RhaS^{L201R}$-His_6. The inserts encoding RhaS and $RhaS^{L201R}$ was generated by PCR with oligos 2349 and 2353. A thrombin cleavage site was primer encoded at the C-terminal end of the protein. The inserts were digested with restriction endonucleases NdeI and XhoI, and then ligated with similarly digested plasmid pET24b to make pSE328 (encodes RhsS-His_6) and pSE329 (encodes $RhaS^{L201R}$-His_6).

β-galactosidase assays. β-galactosidase assays were performed as described previously **[87]**, using the growth protocol of Neidhardt **[88]**and the assay method of Miller **[89]**. Briefly, all strains for β-galactosidase assays were grown in three serial steps: tryptone-yeast extract culture with ampicillin; overnight culture (MOPS-buffered minimal medium containing 0.04% glycerol as a limiting carbon source and ampicillin); and growth culture (MOPS-buffered minimal medium with 0.4% glycerol, 200 μg/ml ampicillin, and with or without 0.4% L-rhamnose). Specific activities were averaged from three independent experiments with two replicates in each experiment. Standard deviation of the mean was calculated for all the variants.

Western blot. Cultures of the strains used in β-galactosidase assays were lysed by boiling at 100°C for 20 minutes, loaded onto a 12% sodium dodecyl sulfate -polyacrylamide gel, electrophoresed, blotted onto a nitrocellulose membrane, blocked with Odyssey® blocking buffer (LICOR, Lincoln, NE), and incubated with primary and secondary antibodies using

standard procedures. The rabbit anti-RhaR and anti-RhaS primary antibodies were custom-made from Cocalico Biologicals (Reamstown, PA). The mouse anti-DnaK primary antibody was purchased from Abcam (Cambridge, MA). The Alexa Fluor® 680-labled anti-rabbit secondary antibody and IRDye® 800-labled anti-mouse secondary antibody were obtained from Life Technologies (Carlsbad, CA) and LI-COR (Lincoln, NE), respectively. The blot was imaged using an Odyssey Infrared Imaging System (LI-COR, Lincoln, NE).

***In vivo* dose-response experiments.** *In vivo* dose-response assays were performed as previously described [81]. For all the assays performed, SE-1 was dissolved in 100% dimethyl sulfoxide (DMSO). To test the effect of SE-1 on RhaS and RhaS variants, cell cultures of RhaS and its variants were grown to an optical density at 600 nm (OD600) of ~0.1 and then mixed with various concentrations of analogues (final concentration of DMSO, 6.2%), induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) (final concentration) and 0.2% L-rhamnose (final concentration) for 3 hours at 37°C, lysed, and β-galactosidase activity was measured. Uninduced (no IPTG and no analogues) and uninhibited (1 mM IPTG, 0.2% L-rhamnose and no analogues) controls were included for each of the *rhaB-lacZ* and control strain, and used to normalize β-galactosidase activity, as described previously [81].

To assay the effects of SE-1 analogues, similar *in vivo* dose-response assay was carried out with some exceptions. Twenty-four commercially available compounds were obtained from Sigma Aldrich (St. Louis, MO), and other analogues were synthetized by Dr. Jeff Aube's group. Bacterial cultures of SME4382 which was a *virB-lacZ* reporter fusion and contained a plasmid expressing VirF, and SME3359 which was a control strain carrying a *lacZ* reporter fusion (*hts-lacZ*) with a synthetic promotor that was repressed by LacI and did not require VirF for activation, were used to test all analogues. Cells were grown the same way as mentioned above,

and then induced with 1 mM IPTG (final concentration). For all the assays, fifty percent inhibitory concentrations (IC_{50}) were calculated and graphs were drawn using Graphpad Prism (GraphPad, La Jolla, CA). All data plotted were from three independent experiments with two replicates in each experiment.

***E. coli* growth rate experiments.** To test whether the SE-1 and its analogues had any impact on the growth of the bacterial strains used for in vivo dose-response assays (SME4382 and SME3359), the growth rate in the presence and absence of the compounds were compared using a procedure described previously[81], except the bacterial cultures were grown in TY broth. Briefly, the cells were grown at 37°C in TY broth plus 1 mM IPTG (final concentration), in 24-well microtiter plate, either with compounds (44 μM, 6.2% DMSO) or with DMSO (6.2%) alone, in a PowerWave XS plate reader (BioTek Instruments).

Overexpression of RhaR-NTD-His_6, RhaR-His_6, and RhaR-DBD-His_6. The C-terminal His_6 tagged RhaR-NTD (residue 1 to 172) was first overexpressed and purified from *E. coli* BL21 (DE3) cells (Novagen) for crystallization screen. Cells were first grown in TY with 50 μg/ml kanamycin at 37°C until cells reached an OD_{600} of 0.6, and then induced with 0.5 mM IPTG (final concentration) and incubated at 15°C overnight. Once crystal condition was found, the RhaR-NTD-His_6 protein was overexpressed in *E. coli* B834 (DE3) cells (Novagen) to incorporate seleno-methionine, to allow solution of the phase problem. To overexpress seleno-methionine incorporated RhaR-NTD-His_6 protein, cells were grown using a protocol modified from New England Biolabs. Briefly, cells were first grown in minimal media supplemented with 50 μg/ml L-methionine and kanamycin at 37°C to an OD_{600} of 0.6, and then pelleted and resuspended in a fresh minimal media containing 50 μg/ml kanamycin but without methionine or seleno-methionine. The cells were grown for 4 hours at 37°C to deplete methionine, and then

incubated with 50 μg/ml seleno-methionine for 30 minutes. The cells were induced with 0.5 mM IPTG (final concentration) and incubated at 25°C overnight.

RhaR-His_6 was overexpressed in the strain BL21 (DE3) (Novagen). The overall overexpression procedure was similar to the procedure overexpressing RhaR-NTD-His6, except that no L-methionine or seleno-methionine was added, and cells were induced with 0.5 mM IPTG once the OD_{600} reached 0.6.

Overexpression of RhaR-DBD-His_6, and His_6-RhaS-DBD. Both of the proteins were overexpressed in the strain BL21 (DE3) (Novagen) using the following procedure. Cells were first grown in TY with 50 μg/ml kanamycin for RhaR-DBD-His6 and 100 μg/ml ampicillin for His_6-RhaS-DBD at 37°C until cells reached an OD_{600} of 0.6. The cells were then induced with 0.5 mM IPTG (final concentration) and incubated at 15°C overnight.

Overexpression of RhaS-His_6 and $RhaS^{L201R}$-His_6 with rhamnose. C-terminal His_6 tagged RhaS and $RhaS^{L201R}$ was overexpressed in strain BL21 (DE3) (Novagen) using the following procedure. Cells were first grown in TY with 100100 μg/ml ampicillin at 37°C until cells reached an OD_{600} of 0.6. The cells were then incubated on ice for 15 minutes. 0.2% rhamnose (finial concentration) and 0.5 mM IPTG (final concentration) were then added to the cells. The cells were then incubated at 15°C overnight.

Overexpression of RhaR fusion protein with chintin-binding domain(CBD) and intein domain at the N-terminus. RhaR was overexpressed as a CBD-intein fusion protein from plasmid pTYB11 (New England Biolabs) in *E.coli* strain ER2566 (New England Biolabs). The protein was overexpressed by using an lactose-driven (instead of IPTG-driven) auto-induction protocol [90]. Briefly, cells were grown in ZYM-5052 media [1% tryptone (w/v), 0.5% yeast

extract (w/v), 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 0.5% glycerol (v/v), 0.05% glucose (w/v), α-lactose (w/v), 2 mM $MgSO_4$, 10 μM $FeCl_3$, 4 μM $CaCl_2$, 2 μM $MnCl_2$, 2 μM $ZnSO_4$, 0.4 μM $CoCl_2$, 0.4 μM $CuCl_2$, 0.4 μM $NiCl_2$, 0.4 μM Na_2MoO_4, 0.4 μM Na_2SeO_3, 0.4 μM H_3BO_3] with 200 μg/mL ampicillin at 37°C until cells reached an OD_{600} of 0.6. The cells were then moved to 25°C until the following morning for induction.

Purification of RhaR-NTD-His_6, RhaR-His_6, RhaR-DBD-His_6, and His_6-RhaS-DBD. The purification of all four proteins follows the procedure described below. Cells were harvested by centrifugation. The cell pellets were re-suspended in binding buffer containing 20 mM Na phosphate pH 7.6, 500 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 5 mM imidazole, 5% (v/v) glycerol, and then lysed by French press. The lysate was centrifuged at 11000 rpm for 30 minutes. Nickel agarose resin (Glodbio) was equilibrated with binding buffer, mixed with the supernatant for an hour with shaking at 4 °C, and then loaded onto a gravity flow column. The nickel resins were washed with 5 column volumes of binding buffer, 10 column volumes of wash buffer (binding buffer but containing 60 mM imidazole). The proteins were eluted with 5 column volumes of elution buffer (binding buffer but containing 500 mM imidazole). The elution fractions with most concentrated proteins were then pooled and loaded onto a HiPrep 16/60 Sephacryl S-300 HR column (GE healthcare), and eluted with buffer containing 10 mM tris-acetate pH7.4, 500mM NaCl, 1mM EDTA, 0.5 mM TCEP, 1 mM dithiothreitol (DTT), 5% glycerol. Protein fractions were pooled and concentrated using Amicon® ultra centrifugal filter unit (Millipore) with a molecular mass cutoff 10 kDa. The final concentrations of RhaR-NTD-His_6, RhaR-His_6, and RhaR-DBD-His_6 were at 6 mg/ml, 2 mg/ml, and 2.5 mg/ml, respectively. His_6-RhaS-DBD protein fractions were also combined, loaded onto HiPrep 26/10 Desalting column, and eluted with buffer containing 100mM K_2HPO_4 pH 6.6, 200

mM KCl, 1 mM DTT, 5% glycerol. Protein fractions with most concentrated proteins were combined and used for further analysis, and the protein final concentration was at 1 mg/ml.

Purification of CBD-intein-RhaR. Cells were harvested by centrifugation. The cell pellets were re-suspended in column buffer with 20 mM HEPES pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.5 mM TCEP, and then lysed by French Press. The lysate was centrifuged at 11000 rpm for 30 minutes. 1 ml of chitin resin (New England Biolabs) was equilibrated with column buffer, mixed with the supernatant for an hour with shaking at 4°C, loaded onto a gravity flow column, and then washed with 20 column volumes of column buffer. 10 column volume of cleavage buffer (column buffer with 100 mM DTT) was added, and then column was gently agitated by shaking for at least 16 hours to promote the intein-mediated cleavage of RhaR from the fusion protein. The untagged RhaR was eluted, and then concentrated with a 4 ml Amicon® ultra centrifugal filter unit (Millipore) with a molecular mass cutoff of 10 kDa.

Purification of RhaS-His_6 and $RhaS^{L201R}$-His_6 with rhamnose. Cells were harvested by centrifugation. The cell pellets were re-suspended in binding buffer containing 20 mM Na phosphate pH 7.6, 500 mM NaCl, 0.2% rhamnose, 0.5 mM TCEP, 5 mM imidazole, 5% (v/v) glycerol, and then lysed by French press. The lysate was centrifuged at 11000 rpm for 30 minutes. Nickel agarose resin (Glodbio) was equilibrated with binding buffer, mixed with the supernatant for an hour with shaking at 4 °C, and then loaded onto a gravity flow column. The nickel resins were washed with 5 column volumes of binding buffer, 10 column volumes of wash buffer (binding buffer but containing 60 mM imidazole). The proteins was eluted with 5 column volumes of elution buffer (binding buffer but containing 500 mM imidazole). The elution fractions with most concentrated proteins were then pooled and loaded onto a HiPrep 16/60 Sephacryl S-300 HR column (GE healthcare), and eluted with buffer containing 10 mM tris-

acetate pH7.4, 500mM NaCl, 1mM EDTA, 0.5 mM TCEP, 1 mM dithiothreitol (DTT), 5% glycerol. Protein fractions were pooled and concentrated using Amicon® ultra centrifugal filter unit (Millipore) with a molecular mass cutoff 10 kDa. The final concentration of RhaS- His_6 and $RhaS^{L201R}$-His_6 were at 2 mg/ml.

Purification and crystallization of ToxT. The expression and purification of ToxT was performed as described previously [38], with a few exceptions. Briefly, ToxT was overexpressed as a ToxT-intein-CBD fusion from plasmid pTXB1 (New England Biolabs), by auto-induction in ZYM-5052 media [90] with 200 µg/mL ampicillin using strain BL21 (DE3) (Novogen) instead of strain BL21-CodonPlus® (DE3)-RIL (Stratagene) as previously described [38]. The initial purification was carried out using a chitin affinity column (New England Biolabs) with gravity flow. The CBD tag was on column removed by 100 mM DTT. The eluent from the chitin column, which contained untagged ToxT, was loaded onto a HiTrap SP Sepharose Fast Flow cation exchange column (GE healthcare) in buffer containing 20 mM Tris-HCl pH 6.8, 33.3 mM DTT, and 50 mM NaCl. Buffer A and B contained 0.05 M and 1 M NaCl, respectively, and the protein was eluted using a gradient from 0 to 100% buffer B and protein peak was found at 88% buffer B. The fractions containing purified ToxT protein were combined and then concentrated to 1.65 mg/mL for crystallization screening using a 4 ml Amicon® ultra centrifugal filter unit (Millipore) with a molecular mass cutoff of 10 kDa. All crystallization screening was conducted in Compact Jr. or Clover Jr. (Rigaku Reagents) sitting drop vapor diffusion plates incubated at 293 K using 0.75 µl of protein and 0.75 µl crystallization solution equilibrated against 75 µl of the latter. Crystals displaying needle (~100 µm × 10 µm) or plate (~60 µm × 20 µm) morphology formed overnight from various screens. Plate shaped crystals, used for data collection, were obtained from the Proplex HT screen (Molecular Dimensions Inc.)

condition H10 [5% (w/v) PEG 4000, 10% (v/v) 2-propanol, 0.1 M MES pH 6.5, 200 mM $MgCl_2$], a condition that shared no components with the crystallization solution used by Lowden *et al* [38]. Crystals were transferred to a fresh drop composed of 80% crystallization solution and 20% ethylene glycol and stored in liquid nitrogen.

Data collection and structure refinement of ToxT. X-ray diffraction data were collected by Kevin Battaile at the Advanced Photon Source beamline 17-ID using a Dectris Pilatus 6M pixel array detector, and solution of the structure was performed by Scott Lovell. Intensities were integrated using XDS [91], the Laue class analysis and data scaling were performed with Aimless[92] which suggested that the highest probability Laue class was 2/m and space group $P2_1$. The Matthew's coefficient [93] (Vm=2.3 Å3/Da, % solvent=46.8) suggested that asymmetric unit contained a single molecule. Structure solution was conducted by molecular replacement with Phaser [94] via the Phenix [95] interface using a previously determined non-isomorphous structure of ToxT (PDB code 3GBG [38]) as the search model. All space groups with 2-point symmetry were tested and the top solution was obtained for a single molecule in the asymmetric unit in the space group $P2_1$. Structure refinement and manual model building were conducted with Phenix and Coot [96] respectively. TLS refinement [97] [98]was incorporated in the latter stages to model anisotropic atomic displacement parameters. Structure validation was conducted with Molprobity [99] and figures were prepared using the CCP4MG package [100]. Refined atomic coordinates and experimental structure factors have been deposited to the Protein Data Bank (PDB code 4MLO).

Crystallization of RhaR-NTD-His_6. The rhamnose-bound RhaR-NTD-His_6 was crystallized in hanging drops using 1 µl of protein and 1 µl well solution which contains 100 mM MES pH 6.0, 50 mM $Ca(OAc)_2$, 12 % PEG 3350. The crystals were soaked in well solution with 50 mM

rhamnose for 10 minutes. After transferring to a fresh drop containing 70% crystallization solution and 30% ethylene glycol (V/V), crystals were flash frozen with liquid nitrogen. The rhamnose-free RhaR-NTD-His_6 was also crystallized in hanging drops with 1 µl of protein and 1 µl well solution which contains 100 mM MES pH 6.0, 50 mM $Ca(OAc)_2$, 6 % PEG 3350. Crystals were transferred to a fresh drop composed of 80% crystallization solution and 20% (v/v) ethylene glycol (RhaR-NTD-EG), glycerol (RhaR-NTD-GLY), PEG 200 or PEG 400 for 1 minute and stored in liquid nitrogen.

X-ray data collection and structure determination of rhamnose-bound RhaR-NTD. X-ray data for the full rhamnose-occupied Met-RhaR-NTD were collected remotely at the Stanford Synchrotron Radiation Lightsource (SSRL), and data for the partial rhamnose-occupied native RhaR-ND were collected remotely at Advanced Photon Source (APS). All data were processed by Haiyan Zhao with the program HKL2000 [101]. The structure was determined at 2.05 Å resolution using the single-wavelength anomalous dispersion method with the program SOLVE/RESOLVE [102]. Automated model building with the program PHENIX [95] resulted in a model covering ~80% of the protein (205 out of 258 residues). This model was used in further model building and refinement with the program COOT [96] and PHENIX.

X-ray data collection and structure determination of rhamnose-free RhaR-NTD. X-ray data for the rhamnose-free RhaR-NTD were collected at the Advanced Photon Source (APS) beamline 17-ID using a Dectris Pilatus 6M pixel array detector by Kevin Battaile. Structures were determined by Scott Lovell. Structure solution was conducted by molecular replacement with Phaser [94] using a previously determined isomorphous structure of RhaR-NTD as the search model. Structure refinement and manual model building were conducted with Phenix [95] and Coot [96] respectively. Disordered side chains were truncated to the point for which

electron density could be observed. TLS refinement [97, 98] was carried out in the latter stages of refinement to model anisotropic atomic displacement parameters. Structure validation was conducted with Molprobity [99] and figures were prepared using the CCP4MG package [100]

Electrophoretic mobility shift assays (EMSA) for RhaR and RhaR-DBD. To determine the active protein concentrations and the dissociation constants (K_d) for RhaR and RhaR-DBD, EMSAs were performed in which either RhaR or RhaR-DBD was mixed with fluorescently labelled DNA in reaction buffer consisting of 10 mM Tris-acetate, pH 7.4, 50 mM KCl, 50 mM L-rhamnose (when present), 1 mM EDTA, 0.05% NP40, 1 mM DTT, and 5% glycerol. The reactions were incubated until the system reached equilibrium, and then loaded onto 6% polyacrylamide gel. The gels (6% acrylamide, polyacrylamide to bis-acrylamide ratio, 60:1, w/w) were soaked in running buffer (22 mM Tris-borate, 5 mM EDTA), and pre-run at 150V for at least 15 minutes prior to loading samples.

Electrophoretic mobility shift assays (EMSA) for RhaS and RhaS-DBD. To determine the active protein concentrations and the dissociation constants (K_d) for RhaS and RhaS-DBD, EMSAs were performed in which either RhaS or RhaS-DBD was mixed with fluorescently labelled DNA in reaction buffer consisting of 10 mM Tris-acetate, pH 7.4, 50 mM KCl, 50 mM L-rhamnose (when present), 1 mM EDTA, 1 mM DTT, and 5% glycerol. The reactions were incubated until the system reached equilibrium, and then loaded onto 6% polyacrylamide gel. The gels (6% acrylamide, polyacrylamide to bis-acrylamide ratio, 60:1, w/w) were soaked in running buffer (22 mM Tris-borate, 5 mM EDTA), and pre-run at 150V for at least 15 minutes prior to loading samples. The gels used to measure full length RhaS DNA binding also contained 0.2% rhamnose during gel casting.

RhaR and RhaS DNA bending detected by EMSA. To determine the angle of RhaR DNA bending in the presence and absence of L-rhamnose, the DNA bending assay was performed as described by Wu & Crothers (1984) [103]. Briefly, RhaR protein was mixed and incubated with each DNA fragment, loaded onto the gels, and electrophoresed for approximately 3 hours. The five 292 bp DNA fragments with the 51 bp RhaR DNA binding site at different positions were amplified and labeled with IR700 fluorescence tag via PCR. To calculate the bend angle, method developed by Thompson & Landy (1988) [104] was used. T-test [105] was performed to analyze the significance of the bend angles measured in the presence and in the absence of L-rhamnose.

To determine the angle of RhaS DNA bending in the presence of L-rhamnose, the DNA bending assay was performed as described above. Briefly, RhaS protein was mixed and incubated with each DNA fragment, loaded onto the gels, and electrophoresed for about 3 hours. Each of the 291 bp DNA fragments with the 51bp RhaS DNA binding site at various positions was amplified and labeled with IR700 fluorescence tag via PCR.

Active protein concentration assay. Total protein concentration was determined by Bradford assays. The concentration of active protein in each protein sample was determined using EMSA assays under stoichiometric conditions – with the DNA concentration in great excess over the protein concentration - as previously described [53]. DNA was mixed with serial dilutions of the protein of interest and an EMSA was performed. A linear relationship between protein concentration and the amount of DNA bound was determined. This linear trend line was used to determine the protein concentration necessary to bind all of the DNA, and the percent active protein was calculated as a function of the total DNA concentration divided by the total concentration of RhaR or RhaS protein, as described previously [106, 107]. Measurements of

active protein concentration were performed very close in time and with the same protein aliquots as the final equilibrium dissociation constant assays to ensure accurate active protein concentration values were used in the dissociation constant calculations.

Electrophoretic mobility shift assays (EMSA) to detect DNA-binding by ToxT. To determine whether purified ToxT protein was active and responsive to effector fatty acid, EMSAs were performed. The 54-bp double stranded DNA for ToxT binding was annealed including the ToxT binding site sequence from the *tcp* promoter region. Fatty acid oleic acid was dissolved in 100% methanol. To carry out binding assay, 1.8 μM of ToxT was mixed with DNA in reaction buffer consisting of 10 mM Tris-acetate, pH 7.4, 50 mM KCl, 0.008% oleic acid (when present), 1 mM EDTA, 1 mM DTT, and 5% glycerol.. The reactions were incubated at 30 °C for 30 minutes, and then loaded onto 6% polyacrylamide gel. The gels (6% acrylamide, polyacrylamide to bis-acrylamide ratio, 60:1, w/w) were soaked in running buffer (22 mM Tris-borate, 5 mM EDTA), and pre-run at 150V for at least 15 minutes prior to loading samples. EMSAs were also performed to test whether SE-1 inhibits DNA-binding by ToxT. SE-1 was dissolved in 100% DMSO. 1.8 μM of ToxT was mixed with SE-1(or DMSO only) first and incubated on ice for 30 minutes; DNA was then added and incubated at 30°C for 30 minutes before loading onto the gel.

Fluorescence-based thermal shift assay. Binding of SE-1 to purified VirF was performed using a thermal shift assay with the fluorescent dye Sypro Orange [108, 109] (Molecular Probes). A protocol of Ericsson *et al.* [109] was used with the following minor modifications. The binding of SE-1 to an unrelated purified protein MBP-NS1-NTD (more than 95% pure) was also assayed as a control. Each reaction mixtures contained 1 μM protein, 20× Sypro Orange, 0.76× EMSA buffer (7.6 mM Tris-acetate [pH 7.4], 0.76 mM K-EDTA, 38 mM KCl, and 0.76 mM

DTT), 80 μM SE-1 with 4% DMSO or 4% DMSO only. Low-profile 0.1-ml PCR 8-tube strips with optically clear flat caps (USA Scientific) were used to carry each reaction. Each reaction was then put in a StepOnePlus real-time PCR thermocycler (Applied Biosystems), with heating from 25 to 99°C in increments of 0.8°C. Boltzmann sigmoidal model was used to analyze all the data collected and to calculate the T_m with GraphPad Prism software (GraphPad). Data points before and after the fluorescence intensity minimum and maximum, respectively, were excluded from fitting. The values for ΔT_m (T_m changes) are averages for three independent experiments with two replicates in each experiment.

Intrinsic tryptophan fluorescence assay. Serial dilutions of SE-1 were prepared with protein buffer containing 100 mM K_2HPO_4 pH 6.6, 200 mM KCl, 1 mM DTT, 5% glycerol. His_6-RhaS-DBD protein was diluted to a final concentration of 0.9 μM with the aforementioned buffer. Reactions were set up by mixing 1 μl of SE-1 dilutions and 2 ml of protein, and kept on ice before using. SE-1-only controls were also set up by mixing 1 μl of SE-1 dilutions and 2 ml of buffer. Each reaction was placed in a 2 ml cuvette, and the fluorescence was monitored using Varian Cary Eclipse Fluorescence Spectrophotometer (Varian Medical Systems, Palo Alto, CA) . Excitation was at 295nm, and the slit width was 5nm for both excitation and emission. Preliminary experiments indicated that SE-1 has significant absorption at the wavelength at which the tryptophan emits photon; therefore this inner filter effect needs to be corrected. To do so, a 1 ml cuvette was filled with SE-1 only controls, and the UV absorption was scanned and recorded from 280 to 380 nm using Beckman UV/Vis spectrophotometer (Beckman Coulter). The corrected fluorescence intensity was calculated as described previously [110].

Protein peptides mapping by LC/MS. The protein peptides mapping by LC/MS experiments were performed by Dr. Nadya Galeva of the University of Kansas Mass Spectrometry/Analytical

Proteomics Laboratory as described previously [111]. Briefly, NanoAcquity chromatographic system (Waters Corp., Milford, MA) and Binary Solvent Manager were used to develop gradient with solvents of highest purity (Optima LC-MS grade, Fisher Scientific) and to deliver analytes through a reverse-phase column (Thermo Scientific Acclaim PepMap300) to an electrospray source of an LTQ-FT mass spectrometer (ThermoFinnigan, Bremen, Germany). A linear gradient was developed in 50 minutes with a flow rate of 10 µl/min. The LTQ-FT mass spectrometer was operated in a data-dependent acquisition mode. Thermo Scientific™ Xcalibur™ software (Waltham, MA) was used to acquire the data. Mascot software [112, 113] was used for database search and protein mapping, setup to search SwissProt or custom protein sequence with the digestion enzyme trypsin. Mass tolerance was 20 ppm for precursor ions and 0.5 u for fragment ions. Carbamidomethyl was set as a fixed modification of cysteine residues, oxidation of methionine residues and acetylation of protein N-terminus were specified as variable modifications. The results were exported into Scaffold software (Proteome, Portland, OR) for statistical analysis and data sharing.

Chapter 3

Identifying Small Molecule Inhibitor Binding Site in RhaS

Previously, using high-throughput screening, the Egan lab has identified a small molecule inhibitor SE-1 that inhibited the activity of many AraC family activators, including RhaS, RhaR, and VirF [80, 81]. It was also found that SE-1 inhibited RhaS-DBD with an IC50 of 10 μM. The inhibition of SE-1 to the RhaS-DBD was to the same extent as full-length RhaR, which suggested that the RhaS N-terminal domain was not required for SE-1 to inhibit RhaS, and the DBD was the target for SE-1 [81]. However, it was unclear where the binding site of SE-1 was in the RhaS DBD, and what the mechanism of action of SE-1 was. In an effort to identify the binding site of SE-1, the Egan lab has performed a computational docking study by using online docking servers SwissDock [114], and BSP-SLIM [115]. Both of the servers require 3-D structure of target protein to calculate the possible binding of small molecule ligand [114, 115]. Therefore, AraC-DBD (PDB:2K9S) [34], MarA (PDB: 1BL0) [8] whose structures have been solved were first used to carry out the docking study. Although crystallization attempts of RhaS, Rns and VirF have not yet been successful, by utilizing online structure prediction server ITASSER, we also obtained the modeled structures of RhaS-DBD, Rns-DBD, and VirF-DBD, and used them to perform docking studies. Both SwissDock and BSP-SLIM have predicted several binding site of SE-1 in each of the protein, and different poses of SE-1 were also predicted at each predicted SE-1 binding site. The top model with the most favorable energy score predicted the binding site of SE-1 was located at a small pocket between the two HTH motifs in all of the protein structures tested (Fig. 6).

Here, I have used fluorescence-based thermal shift assay and intrinsic tryptophan fluorescence assay to test whether SE-1 bind directly to VirF and RhaS, respectively. I

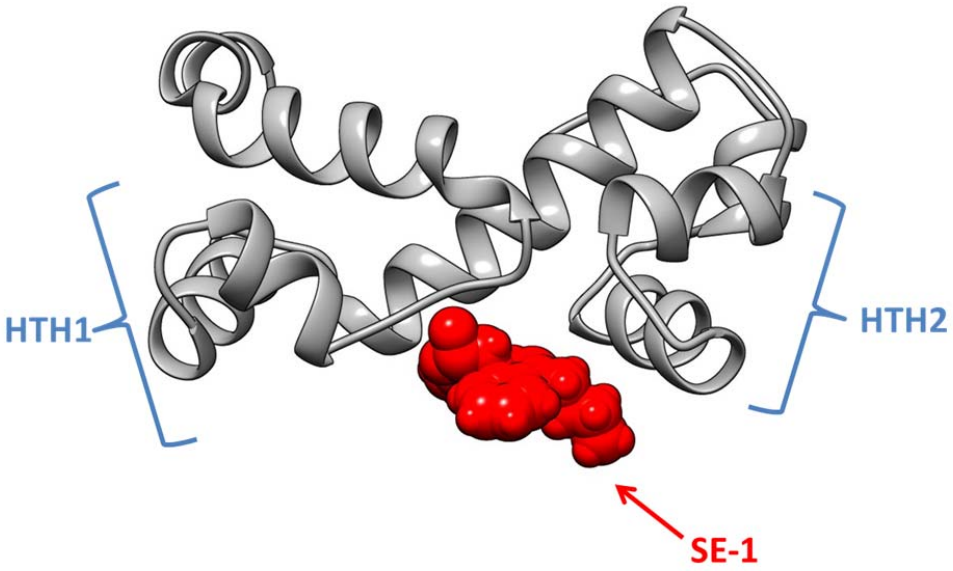


Fig. 6. Top model of SE-1 docked to the RhaS-DBD. The I-TASSER [116] predicted structure of RhaS-DBD is in gray and SE-1 is in red. Only one pose of SE-1 is shown. The helix-turn-helix motif (HTH) 1 and 2 are indicated.

also have used site-directed mutagenesis to test the predicted SE-1 binding site on the RhaS protein.

SE-1 binds directly to VirF. The finding that SE-1 blocked DNA binding by VirF but not by the non-AraC family proteins LacI and CRP, suggested that SE-1 might bind directly to VirF. To test the hypothesis that SE-1 binds to VirF, I performed thermal shift assays [108, 109] using the dye Sypro Orange (Molecular Probes, Invitrogen). In this assay, as the temperature increases and the protein begins to unfold, an increased fluorescence signal will be observed upon the binding of Sypro Orange to the exposed hydrophobic residues of the protein. Therefore, the T_m of the protein in the absence and presence of a ligand can be measured, and any melting temperature changes (ΔT_m) of the protein due to the increased protein stability caused by ligand binding can be calculated. Using the thermal shift assay, It was found that the addition of 80 μM SE-1 increased the T_m of MBP-VirF (1 μM) by 0.61°C (±0.09) (Fig. 7A). Although this Tm change seems to be a relatively small, the following evidence supports the conclusion that this likely indicates binding of SE-1 to VirF. Pantoliano *et al.* [108] has tested over 100 different proteins and used a ΔT_m cutoff of 0.5°C or higher to determine the binding of the small molecule ligands to the target proteins. Furthermore, the ΔT_m for VirF with 80 μM SE-1 is much higher than that found with 40 μM SE-1 (0.15°C ± 0.04 [data not shown]). SE-1 is quite hydrophobic and dissolves poorly in water (~1.4 mM). Preliminary tests suggested that fluorescence dye Sypro orange could bind to SE-1 and increase the base level of the fluorescence during the assay. When mixing 160 μM of SE-1 (or higher) with the protein, the fluorescence signal of the protein melting cannot be easily distinguished from the elevated background fluorescence signal.

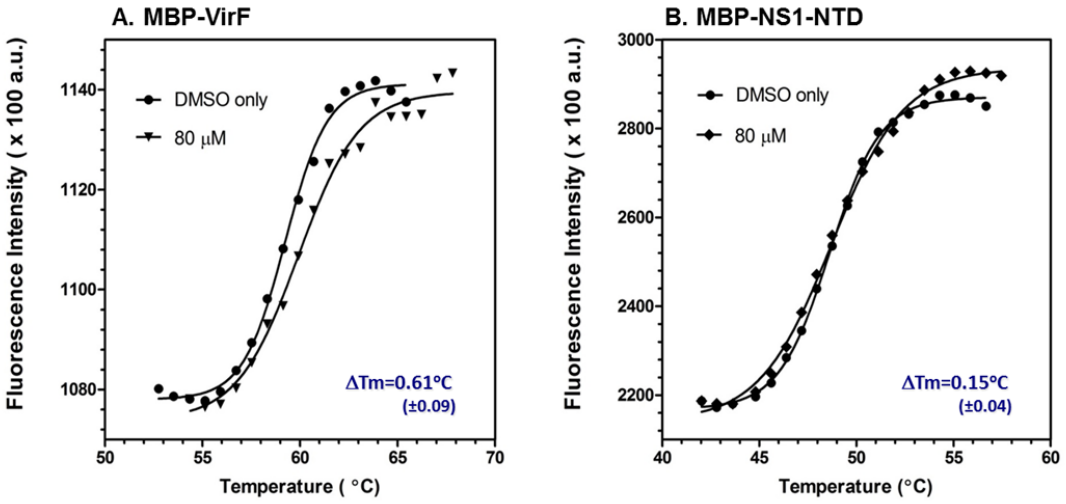


Fig. 7. Thermal shift assay showing that SE-1 binds to VirF protein. Melting curve of MBP-VirF (A) and MBP-NS1-NTD (B). Binding assays were performed in the absence (DMSO only) or presence of 80 mM SE-1. Results of a single experiment (representative of three independent assays with two replicates each) are shown.

Therefore, only 80 μM of SE-1 was tested in the assay. It is likely that the binding saturation of SE-1 to VirF was not achieved at 80 μM, and that a higher ΔT_m would be observed if higher concentrations of SE-1 was tested. Lastly, multiple reports with small T_m changes have been shown to represent true binding of small molecules to proteins [108, 117-119]. Overall, the finding that the binding of 80 μM SE-1 causes T_m changes of VirF by 0.61°C (±0.09) suggested that SE-1 directly binds to VirF.

In order to rule out the possibility that SE-1 bound to MBP since the VirF protein tested was fused with MBP at the N-terminus, I also tested an unrelated MBP fusion protein (MBP-NS1-NTD). NS1 is a poxvirus protein that does not belong to the AraC family. MBP-NS1-NTD protein showed only a 0.15°C (±0.06) change in the T_m (Fig. 7B), and the melting curve of SE-1 binding to the protein almost overlapped with the DMSO only control, suggesting that SE-1 did not bind to MBP. Therefore, we concluded that SE-1 binds directly to VirF.

SE-1 binds directly to RhaS DBD. The finding that SE-1 blocked DNA binding by RhaS-DBD suggested that SE-1 might bind directly to RhaS-DBD. To test the hypothesis that SE-1 binds to RhaS-DBD, I performed intrinsic tryptophan fluorescence assay using purified RhaS-DBD protein. An important feature of intrinsic tryptophan fluorescence is the high sensitivity of tryptophan to its local environment. Changes in the emission spectra of tryptophan often occur in response to protein conformational changes. If SE-1 directly binds to RhaS-DBD, the binding would change the conformation of the protein and thus change the local environment of tryptophans, and I would expect to see a quench in the fluorescence signal upon SE-1 addition. Using this assay, I found that there was a concentration dependent quenching of RhaS fluorescence upon addition of SE-1 (after inner filter correction [110]) that saturates at ~7% reduction in the fluorescence peak (Fig. 8). Both structural and biochemical studies have shown

that L-arabinose binds to AraC [33, 42, 120], despite that the binding of L-arabinose to the AraC-NTD only induced 5 - 10% reduction in the total fluorescence emission [120]. The small change in fluorescence of arabinose binding to AraC was similar to that of the binding of SE-1 to RhaS-DBD, supporting that SE-1 directly binds to RhaS-DBD.

Mutagenesis to identify SE-1 binding site in RhaS. Computational programs SwissDock and BSP-SLIM predicted that SE-1 bound to a small pocket between the two HTH motifs. To test this prediction, I made substitutions at each of the positions in RhaS that were predicted to be at the pocket, and these position included residues V188, W190, D191, L220, N221, R224, F248, S249, H253, and L257. For each of the positons, I constructed both alanine and arginine substitutions (for positon W190, R224, F248, S249, and H253, only alanine substitution was made) and the rationale for making both substitutions is outlined below. Alanine substitution eliminates the side-chain beyond the β carbon and its non-bulk nature does not change the main-chain conformation nor does it impose any electrostatic or steric effects [121]. If one residue is involved in interacting with SE-1, an alanine substitution will no longer be able to make contact with SE-1 and thus will disrupt SE-1 binding. If several residues are involved in contacting SE-1, making alanine substitution at one position may not have enough impact to disrupt the binding of SE-1. While making an arginine substitution will introduce a large side-chain potentially occupying the space for SE-1 to bind. Since all the variants that I constructed are at the DNA-binding interface, it is expected that some of them would be defective for DNA binding. Therefore, I started by measuring the *in vivo* transcription activation activity of all the variants relative to the wild-type RhaS by β-galactosidase assay. These assays did not involve the addition of SE-1. I found that the activities of many of the variants including V188A, D191A,

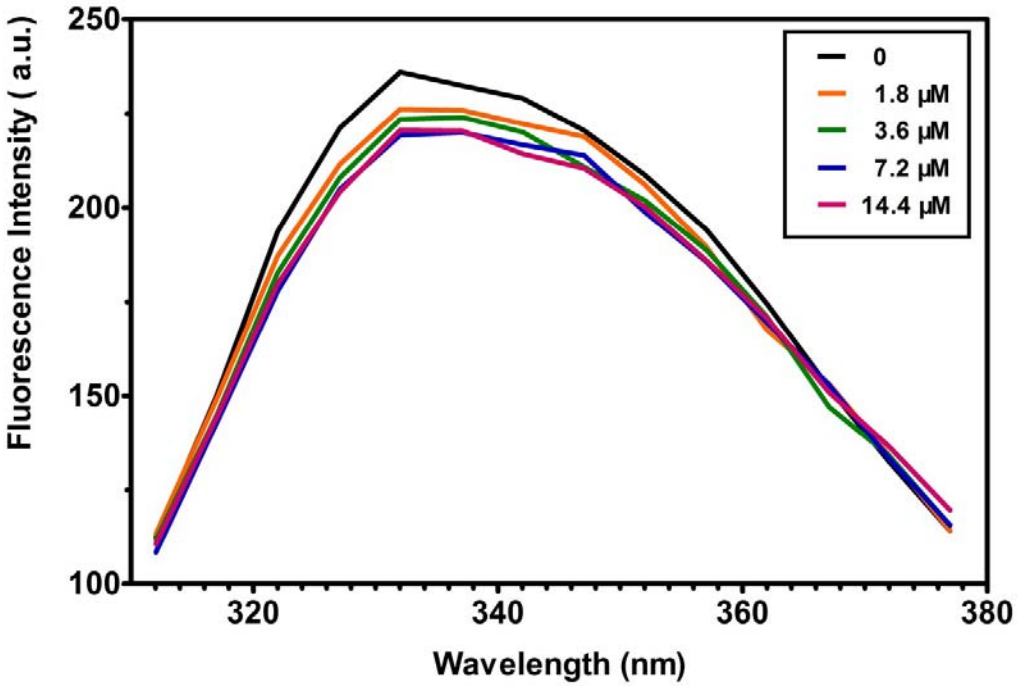


Fig. 8. Quenching of intrinsic RhaS-DBD tryptophan fluorescence by SE-1. Protein was at 0.9 μM. SE-1 concentrations were from 1.8 mM to 14.4 μM, with serial 2-fold dilutions.

D191R, L220A, L220R, N221A, N221R, S249R, H253A, L257A, and L257R was high enough (at least 3% of the wild-type activity) to be further tested in the whole cell based assay with SE-1, and few of the variants including V188R, W190A, R224A, and F248A had very low activity so they would not be assayed in the whole cell based dose-response assay (Table 7). Therefore, I assayed variants V188A, D191A, D191R, L220A, L220R, N221A, N221R, S249R, H253A, L257A, and L257R in the RhaS-activated *rhaB-lacZ* fusion assay *in vivo* in *E. coli* (Fig. 9a), and found that SE-1 inhibited variants V188A, L220A, L220R, N221A, N221R, and H253A to approximate the same extent as WT RhaS, which suggested that these variants did not affect the inhibitory effect of SE-1 to RhaS. The IC_{50} of SE-1 to the wild-type RhaS was about 20 μM, and the IC_{50} values of SE-1 for these variants were approximately 20-30 μM. the IC_{50}s of SE-1 to variants D191A, D191R, S249R, L257A, and L257R were in the range of 60-90 μM (Fig. 9B), which suggested that these variants were less sensitive than wild-type RhaS to inhibition by SE-1 in the dose-response assay. Thus, it was likely that residues D191, S249, and L257 were at or near the SE-1 binding site on RhaS-DBD, which supports the prediction of docking that SE-1 is likely bound to a small pocket between the two HTH motifs. Sequence alignment showed that residue D191 shared sequence identity with RhaR and residue S249 shared sequence identify with Rns and VirF (Fig. 10); however, these two residue do not share any sequence identity with other proteins we analyzed, and moreover, residue L257 does not share any sequence identify with any of the proteins tested, which might contribute to the IC50 differences in SE-1 binding to RhaS, RhaR, VirF, Rns and ToxT. The in vivo transcription activities of variants W190A, R224A, F248A are too low to be assayed with SE-1 in the whole cell based dose-response assay, therefore an assay that does not rely on DNA binding would be needed to assess the effect of SE-1 to these variants. Thermal melt assay measures binding by the Tm change upon the addition of

Variants	% WT Activation	
	(-)L-rhamnose	(+) L-rhamnose
V188A	11	3
V188R	0.3	1
W190A	1	1
D191A	69	244
D191R	103	453
L220A	9	3
L220R	5	3
N221A	31	9
N221R	44	39
R224A	0.01	1
F248A	0.1	1
S249R	21	80
H253A	10	4
L257A	97	301
L257R	89	124

Table 7. Activation activities of variants relative to wild-type RhaS by β-galactosidase assay. β-galactosidase activity was assayed from a single-copy Φ(*rhaB-lacZ*)Δ84 fusion, in a strain with *ΔrhaS*::*kan recA*::*cat* (SME1088) and wt or RhaS variants expressed from a plasmid. Cultures were grown with or without L-rhamnose. Variants were assayed in groups with a wild type RhaS activity range of 0.61-0.72 Miller Units (-) L-rhamnose and 171-180 Miller Units (+) L-rhamnose. Percent of wild-type activation was calculated by dividing the activity of each variant, either (-) or (+) L-rhamnose, by the wild-type value (-) or (+) L-rhamnose, respectively, and multiplying by 100.

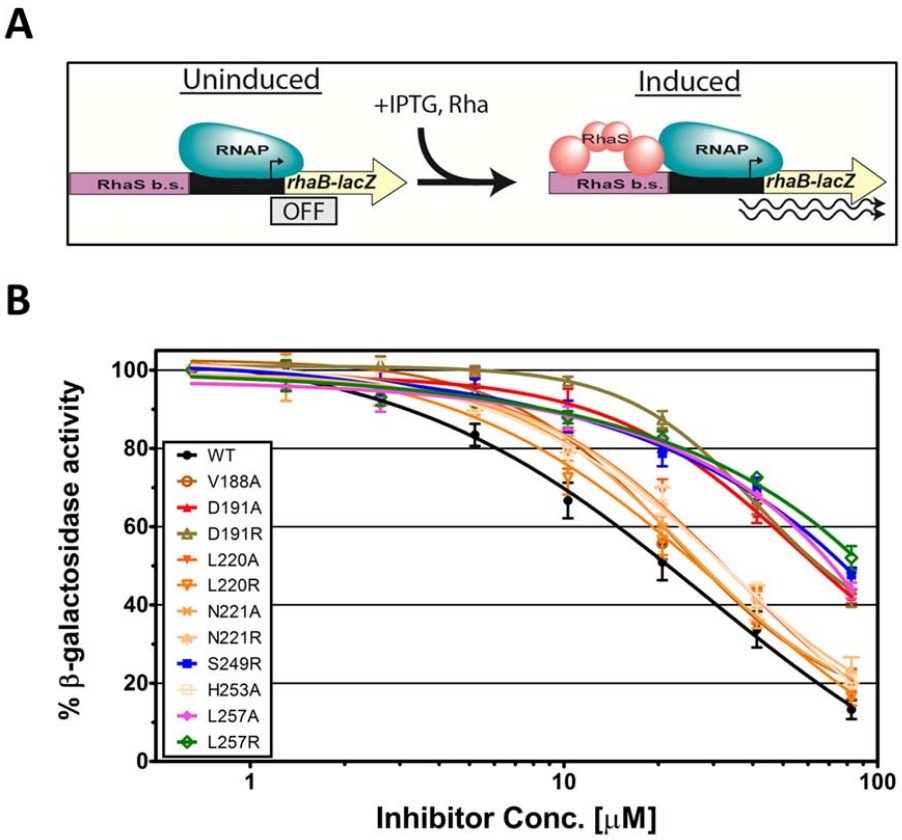


Fig. 9. Reporter fusion to test the effect of SE-1 *in vivo* and inhibition of *in vivo* RhaS variants in *E. coli*. (A) RhaS-activated *rhaB-lacZ* fusion, shown in uninduced, (-) rhamnose state (left) and induced (+) rhamnose state (right). (B) *In vivo* inhibition by SE-1 of β-galactosidase expression from the *rhaB-lacZ* fusion by wild-type RhaS or RhaS variants. Activity in the absence of SE-1was set to 100% for the reporter fusion. Results are the average of three independent experiments with two replicates each.

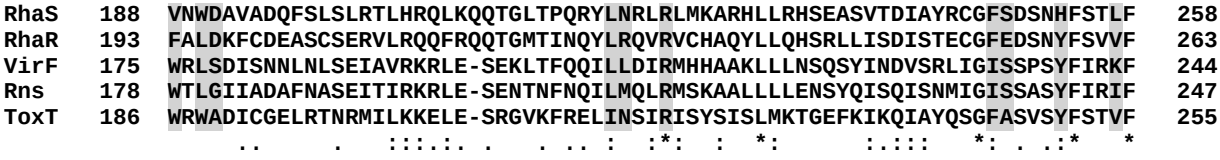


Fig. 10. Sequence alignments of predicted SE-1 binding site in RhaS, RhaR VirF, Rns, and ToxT. Residues predicted to contact SE-1 are highlighted in gray.

SE-1, and intrinsic fluorescence assay measures binding by the fluorescence signal quenching upon SE-1 addition. Therefore both assays could serve alternate methods to test the effect of SE-1 to variants W190A, R224A, F248A.

Chapter 4

Crystal structure of ToxT from *Vibrio cholerae*

ToxT is an AraC family transcriptional activator of *Vibrio cholera* virulence gene expression. ToxT contains a C-terminal DNA binding domain and an N-terminal domain responsible for dimerization and effector binding [38]. As a master virulence regulator, ToxT directly activates the expression of the genes encoding the toxin-coregulated pilus (TCP) and the cholera toxin (CT), and auto-regulate its own expression from the *tcp* promoter [19-23]. TCP is essential for colonization of the human intestine, and CT is the cause of the diarrheal disease that is characteristic of cholera [19-21]. In *V. cholera,* both bile and individual unsaturated fatty acids found in bile has been shown to inhibit ToxT-dependent gene activation [39, 40]. The full-length structure of ToxT determined by Lowden *et al*. had the fatty acid cis-palmitoleic acid (PAM) bound to the N-terminal domain [38]. Although oleic acid is likely the physiological effector of ToxT given its high concentration in bile, both PAM and oleic acid were shown to reduce activation of *tcp* and *ctx* in vivo, and to reduce DNA binding by ToxT *in vitro* [40].

The small molecule inhibitor SE-1 has been previously shown to effectively inhibit transcription activation by several AraC family proteins, including RhaS, RhaR, and VirF [80, 81]. Given that AraC family proteins have a structurally conserved DNA binding domain, we hypothesized that SE-1 might inhibit DNA binding by AraC family proteins in addition to RhaS, RhaR, and VirF, for instance, ToxT. If this was true, we could set up crystallization trials to co-crystalize SE-1 with ToxT since ToxT was stable and the structure of ToxT has been solved by Lowden *et al.* [38], and therefore the chance of getting SE-1 crystalized with ToxT was higher than with the proteins that had never been successfully crystalized. This would provide direct and detailed information about the binding site of SE-1 on AraC family proteins. Here, I have tested whether SE-1 inhibited *in vitro* DNA binding by ToxT, and set up crystallization trials to co-crystalize SE-1 with ToxT.

SE-1 inhibition of *in vitro* DNA binding by ToxT. Electrophoretic Mobility Shift Assays (EMSAs) were performed to investigate whether SE-1 inhibited DNA binding by purified ToxT protein. Protein was purified using methods described previously [38]. Briefly, the ToxT protein was first expressed as a chitin-binding domain -intein fusion protein, and purified through chitin affinity chromatography. The chitin binding-domain was removed by DTT without any extra amino acids at the C-terminus (Fig. 11A). Before testing whether SE-1 inhibited DNA binding by ToxT protein, it was important to know that whether the purified ToxT was active. Therefore EMSAs were carried out to test if purified ToxT can bind to DNA. Since in the presence of fatty acid, ToxT protein is in its inactive form and cannot bind DNA, we also would want to know if purified ToxT protein is responsive to fatty acid. ToxT protein was mixed with 0.008% oleic acid and DNA (or mixed only with methanol and DNA). The results suggested that 0.008% oleic acid was able to fully inhibited DNA binding by ToxT, which was consistent with what Lowden *et al.* has demonstrated [38], and adding methanol to the protein did not affect DNA-binding by ToxT (Fig. 11B). To test if SE-1 inhibits DNA binding by ToxT, protein was mixed with SE-1 (when present) and DNA. Our results indicate that SE-1 was able to inhibit ~90% DNA binding by ToxT at a concentration of 1.34 mM (Fig. 11C). Previously we have shown that SE-1was able to fully inhibit DNA binding by RhaS and VirF at 0.67 mM [80, 81]. However, at 0.67 mM, SE-1 was only able to inhibit ~80% DNA binding by ToxT, and ~90% DNA binding even at 1.34 mM. The concentration of SE-1 required to fully inhibit DNA binding by ToxT would be much higher than the concentration tested to fully inhibit DNA binding by RhaS and VirF, suggesting that although SE-1 could block DNA binding by ToxT, the binding affinity of SE-1 to ToxT may be much weaker than to other proteins tested including RhaS, RhaR, and VirF.

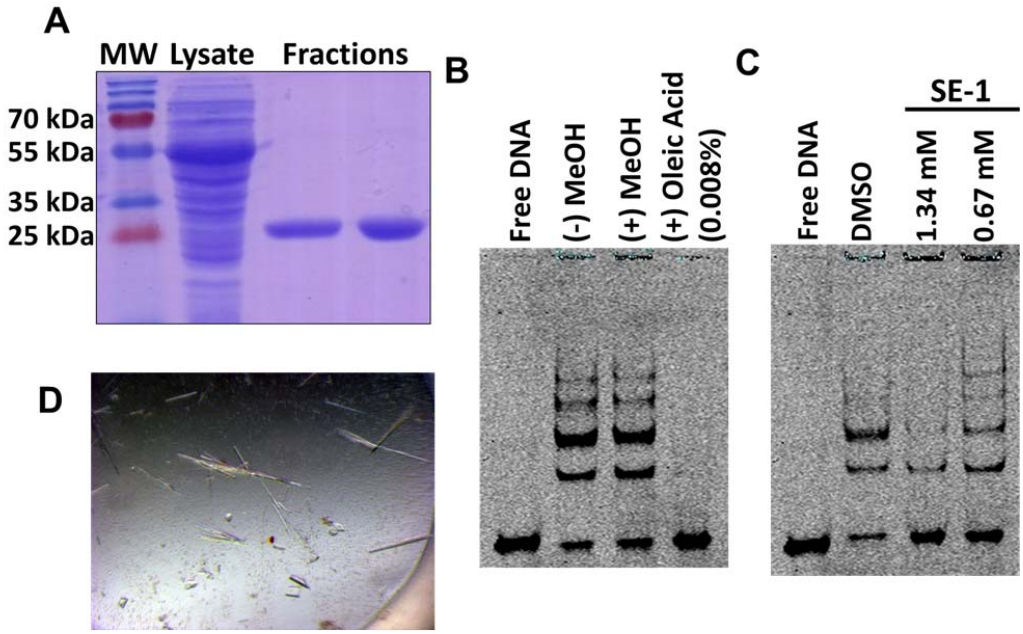


Fig. 11. Purified ToxT protein tested with oleic acid and SE-1 in DNA binding assay, and cystals of ToxT. (A) Purified ToxT protein (12% SDS-PAGE gel). Pre-stained protein ladder (Thermo Fisher Scientific) is in the first lane. ToxT protein was expressed as chitin-binding domain-intein fusion protein (molecular mass: ~ 55 kDa), the chitin-binding domain was then cleaved off by DTT on column, and ToxT protein was eluted without any extra amino acid at C-terminal. (B) EMSA showing ToxT effector oleic acid (in methanol, MeOH) inhibited *in vitro* DNA binding by ToxT. MeOH control was including showing that it did not affect DNA bingind by ToxT. (C) EMSA showing SE-1 inhibited (in DMSO) *in vitro* DNA binding by ToxT. DMSO only control was included showing that it did not affect DNA binding by ToxT. (D) Crystals of ToxT forming either needles (~100 µm x 10 µm) or plates (~60 µm x 20 µm).

Crystallization of ToxT. Once we had shown that SE-1 inhibited *in vitro* DNA binding by ToxT, we set up crystallization trials in hopes of obtaining co-crystals of SE-1 with ToxT, with the goal of obtaining a detailed model of the positon of SE-1 binding to ToxT. As the first attempt to achieve this goal, I initially set up crystallization trials of ToxT using the crystallization condition that Lowden *et al.* (2010) used to successfully crystalize ToxT. However, we weren't able to obtain any crystal of ToxT. Therefore, we screened over 480 different crystallization conditions, and eventually obtained crystals of ToxT with a condition containing 5% (w/v) PEG 4000, 10% (v/v) 2-propanol, 0.1 M MES pH 6.5, 200 mM $MgCl_2$, a solution that is completely different from the published one. The structure was solved to 1.65Å resolution (Fig. 11D).

1.65 Å resolution structure of ToxT. The final refinement and model statistics of ToxT are given in Table 8. The final model of ToxT could be traced in the electron density maps from residues K5 to G272, except for a single disordered residue, G133, which is located in a loop connecting helix α2 to α3 (Fig. 12A). Electron density consistent with PAM was also present (Fig. 12B), as was observed in the original ToxT structure (PDB code 3GBG [38]), although PAM was not added in either case but was acquired from the expression host. Interestingly, residues D101 to E110 could be modelled in this structure, which included helix α1 and a loop region that connects this helix to the β9 sheet. This helix can be thought of as containing two segments which we refer to as α1 and α1' to be consistent with the prior secondary structure assignment for 3GBG [38](Fig. 12C). In addition, three chloride ions were modelled in the C-terminal region of ToxT, which were assigned based on the coordination distances (~3.1-3.3 Å) to neighbouring residues and water molecules. When water molecules were assigned to the

Table 8 Data collection and refinement statistics for ToxT structure

Data Collection	
Unit-cell parameters (Å, °)	a=47.34, b=39.41, c=80.24, β=97.94
Space group	$P2_1$
Resolution (Å)[1]	39.73-1.65 (1.68-1.65)
Wavelength (Å)	1.0000
Temperature (K)	100
Observed reflections	117,532
Unique reflections	35,493
$\langle I/\sigma(I)\rangle$[1]	10.3 (1.9)
Completeness (%)[1]	99.6 (99.8)
Multiplicity[1]	3.3 (3.4)
R_{merge} (%)[1, 2]	8.1 (68.0)
R_{meas} (%)[1, 4]	9.7 (82.6)
R_{pim} (%)[1, 4]	5.2 (43.1)
$CC_{1/2}$ [1, 5]	0.997 (0.714)
Refinement	
Resolution (Å)	39.74-1.65
Reflections (working/test)	33,700/1,777
R_{factor} / R_{free} (%)[3]	16.8/19.4
No. of atoms (Protein/Chloride/PAM/Water)	2,181/3/18/179
Model Quality	
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	0.947
Average B-factor (Å^2)	
All Atoms	25.9
Protein	25.5
Chloride	16.6
PAM	28.1
Water	30.7
Coordinate error(maximum likelihood) (Å)	0.17
Ramachandran Plot	
Most favored (%)	99.6
Additionally allowed (%)	0.4

1) Values in parenthesis are for the highest resolution shell.

2) $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the ith reflection and $\langle I(hkl)\rangle$ is the average intensity of all reflections with indices hkl.

3) $R_{factor} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$; Rfree is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

4) R_{meas} = redundancy-independent (multiplicity-weighted) R_{merge} [92, 122]. R_{pim} = precision-indicating (multiplicity-weighted) R_{merge} [123, 124].

5) $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data [125, 126].

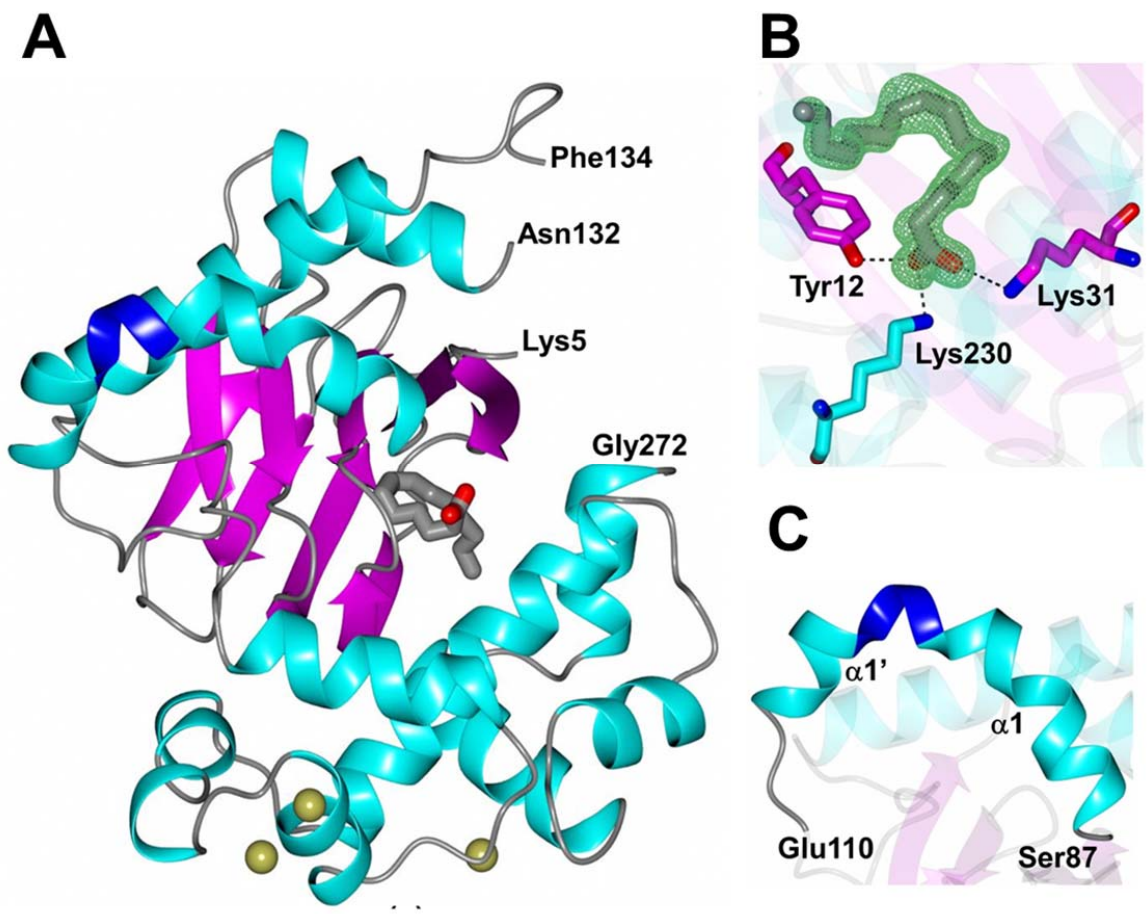


Fig. 12. (A) Asymmetric unit of ToxT (PDB code 4MLO) colored by secondary structure. The N- and C-terminal residues (K5 and G272) for the model are indicated along with the disordered region between AsnN132-F134. The 3_{10} helix spanning L99-D101 is coloured blue. The PAM molecule and chloride ions are drawn as cylinders and gold spheres respectively. (B*) Fo-Fc* omit map contoured at 3σ (green mesh) for PAM and associated hydrogen bonds (dashed lines) to ToxT. (C) Zoomed in view of the region from S87 to E110. Helix α1 spans S87 to I98 and contains a kink at L94. This is followed by a 3_{10} helix spanning L99-D101 and shorter helix from L102 to L107 referred to as α1'.

chloride sites, positive electron density was observed following refinement indicating an underestimation of electrons. Therefore, the modelling of chloride ions at these sites was consistent with the observed electron density and coordination.

The overall structure is similar to PDB code 3GBG reported by Lowden *et al.* (2010) with an RMSD between Cα atoms of 1.00 Å (Lys5 to Gly272) as determined using a Secondary Structure Matching (SSM) [127] algorithm with Superpose via the CCP4 [128] interface. However, there are also differences between the two structures as shown in the per-residue RMSD plot in Fig. 13A and the superimposed structures in Fig. 13B. Specifically, the region between α1 and β9, which was disordered in 3GBG [38] from residues D101 to E110, could be fully traced in the current structure (Fig. 14A). In this region, helix α1 spans S87 to I98 and contains a kink at L94. This is followed by a 3_{10} helix spanning L99-D101 that continues into a shorter helix from L102 to L107 (α1'). Residues T108 to D113 form a connecting loop between α1' and β9. This region appears to be stabilized by residues E156, N160 and I162 from helix α3, which form hydrogen bonding interactions with residues R105 and S109 of the loop region (Fig. 14B). The loop region connecting helices α3-α4 also shows conformational differences relative to 3GBG [38], as depicted in Fig 15A, potentially due to the interactions between residues in the previously disordered region and residues in helix α3. Interestingly, the residue D101 to E110 region is folded over the loop that connects helices α3-α4, spanning K158 to A170, and thus is located sequentially after it. A very similar arrangement can be observed in the structure of the regulatory domain of ExsA where the loop connecting α1 and β9 folds over helix α4 (PDB code 4ZUA [58]). ExsA is an AraC family transcriptional activator that regulates type three secretion system genes in *Pseudomonas aeruginosa* [58, 129]. Residue G133 in 3GBG [38] was ordered, and stabilized by residue K4 through hydrogen bonding interaction. However, both G133 and

K4 were missing from current structure. It is likely that the slight conformational change in the connecting loop region (N132 to D141) disrupted the hydrogen bonding interaction between G133 and K4, causing both residues to become flexible and untraceable in the current structure.

Further analysis was conducted to gauge the quality of fit of the models to the electron density. Analysis of the map-model correlation coefficients via Phenix revealed several regions in 3GBG [38] that display low correlation to the 2*Fo-Fc* map, including the α3-α4 (K158 to A170) loop as shown in Fig. 15B. Although the K158 to A170 loop region was modelled in the 3GBG [38] structure, it is poorly defined, making it difficult to discern the exact position of the residues in this region. By contrast, the electron density in the current structure was clearly traceable in this region, which is reflected in the high correlation coefficient. It should be noted that none of the residues in this loop form hydrogen bond contacts with symmetry related molecules, which suggests that crystal packing is not a factor in the conformational differences relative to 3GBG [38]. Additional differences between the two structures were observed in the loop connecting helices α2-α3 (Asn132 to Asp141) and in part of helix α2 (E122 to V126) (Fig. 15B).

SE-1-ToxT crystallization trials. Co-crystallization trials were set up in attempt to obtain crystals of ToxT in complex with SE-1, and two methods of co-crystallization were used. With the first methods, ToxT protein was mixed with 1.2-fold molar excess of SE-1, and set up crystal screening looking for new conditions that could yield crystals of ToxT with SE-1. I was able to obtain several crystals under new conditions; however, after the X-ray diffraction and structure solving, we found that SE-1 was not in the structure. The second method utilized ToxT crystals that were obtained using condition mention above, and then soak the crystal in SE-1

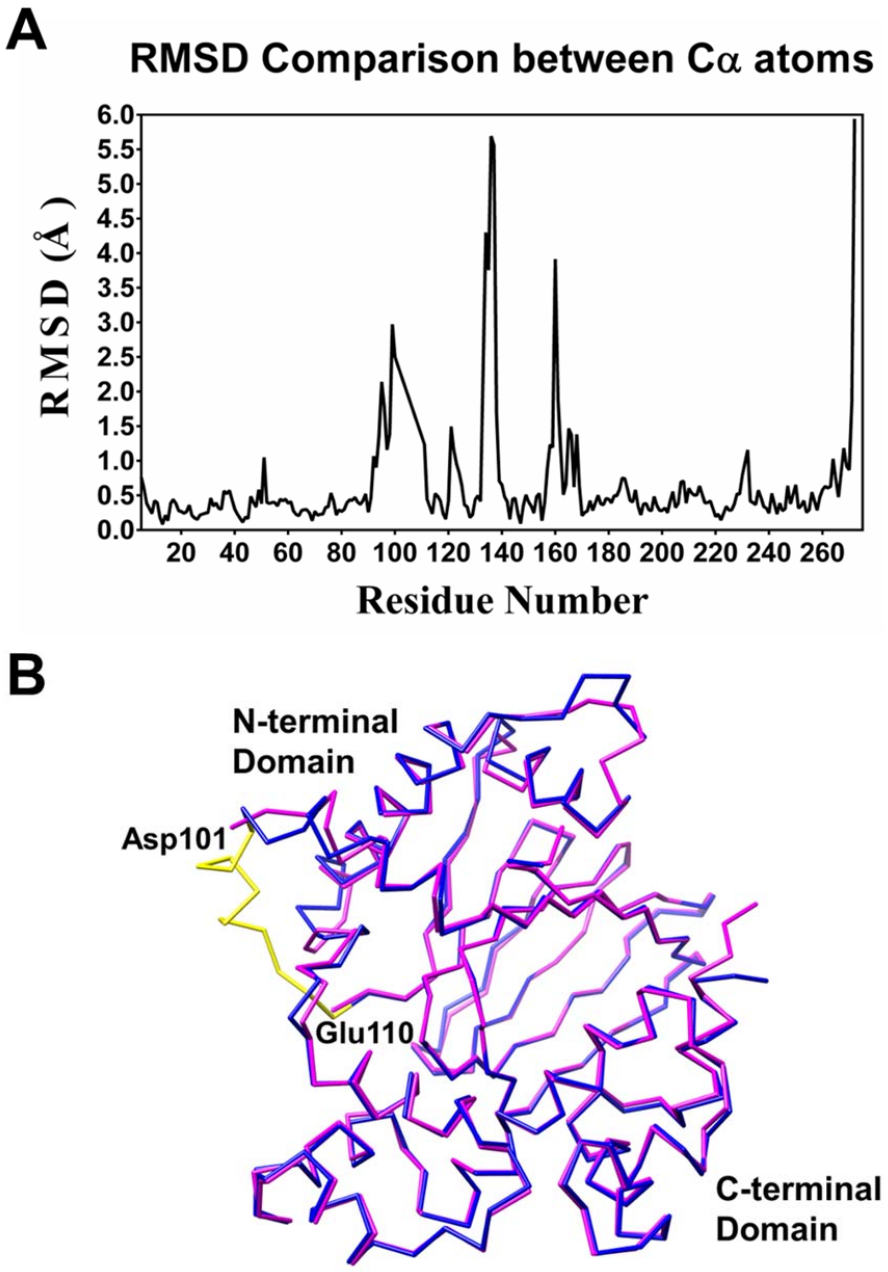


Fig. 13. (A) Plot of RMSD deviations per residue between Cα atoms for ToxT (PDB code 4MLO) and previously determined structure (PDB code 3GBG [38]). (B) Superpostion of (PDB code 4MLO, blue) with the previously determined structure (PDB code 3GBG [38], magenta). The previously disorder region from D101 to E110 is in yellow.

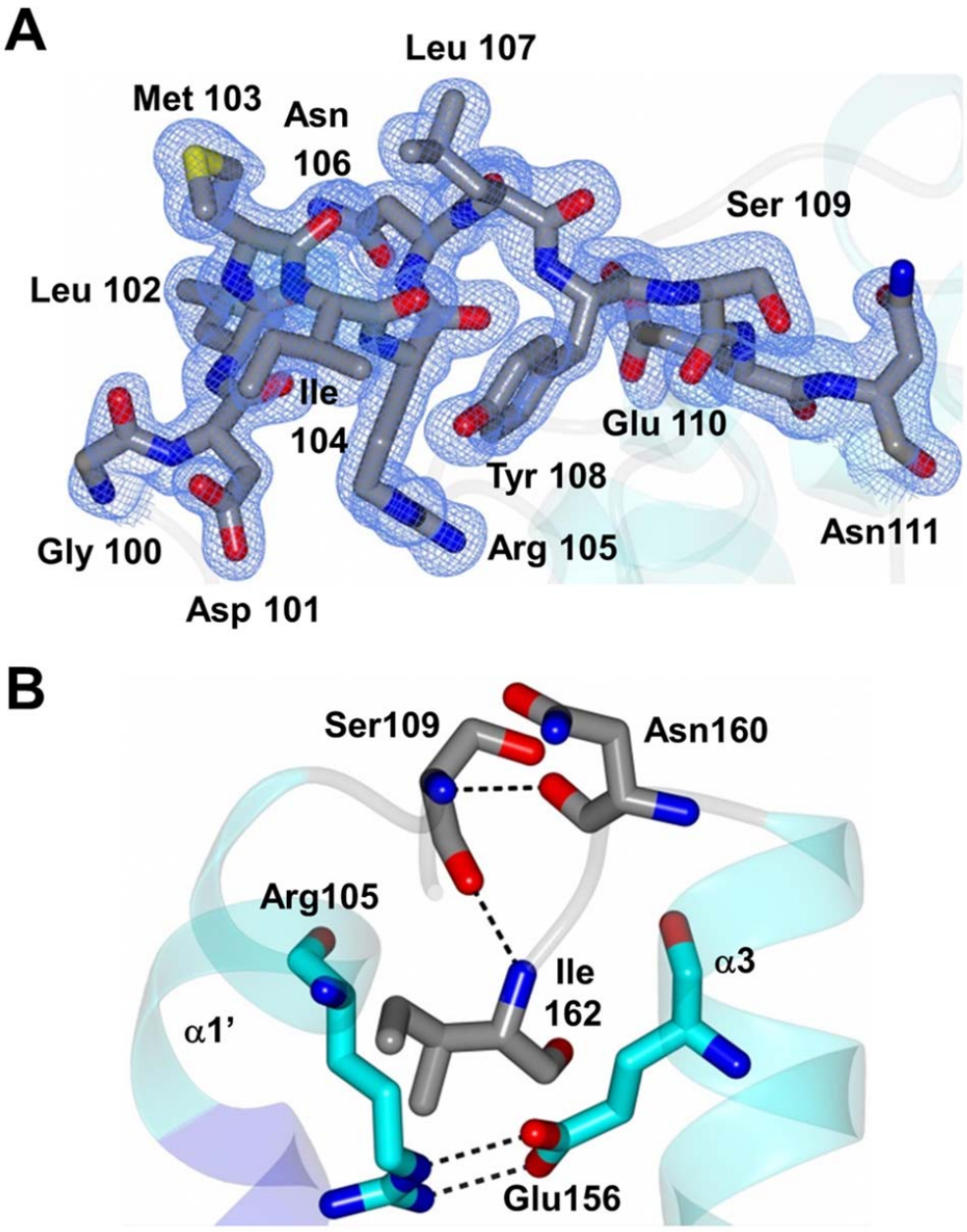


Fig. 14. Loop region between α1' and β9. (A) 2*Fo-Fc* map contoured at 1σ (blue mesh) for residues G100 to N111 which were disordered in PDB code 3GBG [38]. (B). Interactions between α1' and α3. Residues within the α1' (R105) and α3 (E156) helices are coloured cyan. The residues in the loop regions of these helices (S109, N160 and I162) are coloured gray.

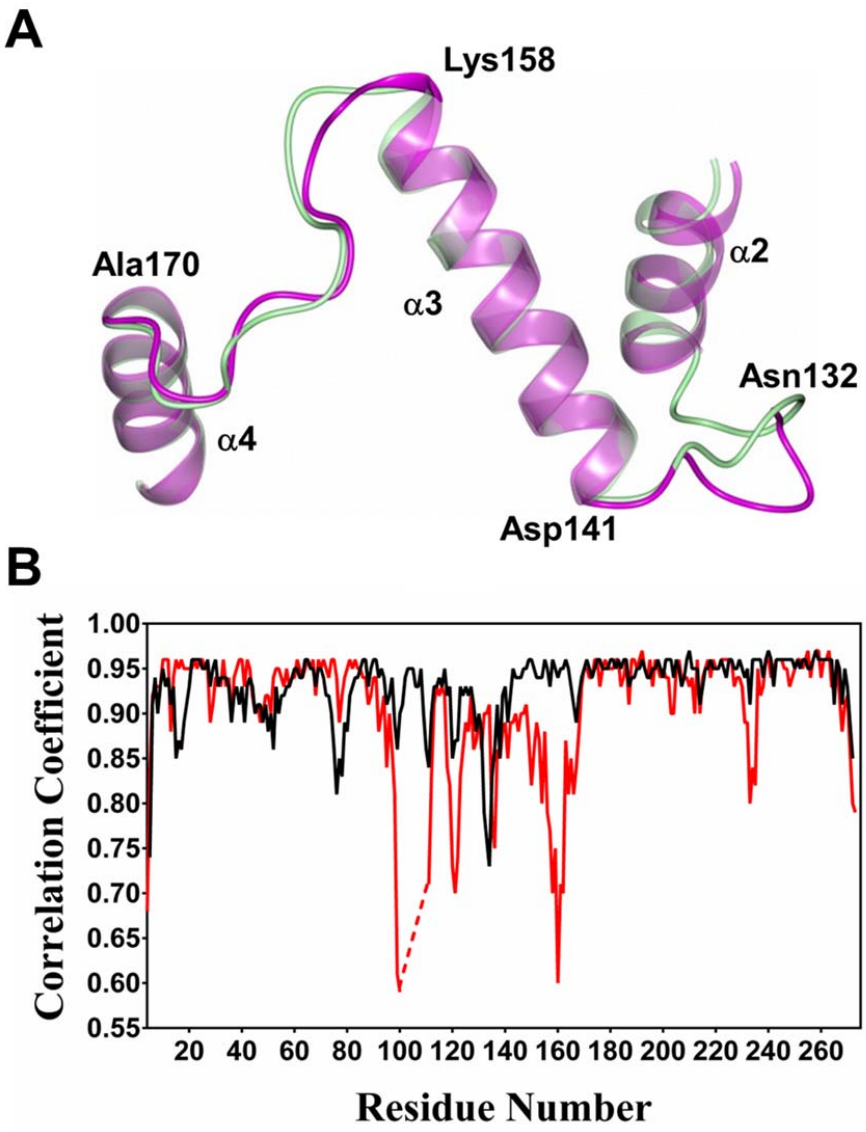


Fig. 15. (A) Comparison of the regions connecting helices α2-α3 and α3-α4 for ToxT (PDB code 4MLO, magenta) with the previously determined structure (PDB code 3GBG [38], green). (B) Comparison of map correlation coefficients (2*Fo*-*Fc*) for ToxT PDB

stock for 20 minutes. Unfortunately, after the X-ray diffraction and structure solving, again, SE-1 was not found in the structure.

Chapter 5

Optimization of the activity of the SE-1 inhibitor

As mentioned in the previous chapters, the small molecule inhibitor SE-1 has been shown to effectively inhibit transcription activation by several AraC family proteins including RhaS, RhaR, VirF, and ToxT [80, 81]. Therefore, SE-1 has the potential to be developed into a novel antibacterial agent. However, SE-1 has shown ~30% inhibition of the control strain at higher concentration, suggesting nonspecific inhibition at higher concentrations, and its potency (IC_{50} in μM range) was not high enough for therapeutic applications. Thus, optimization of SE-1 would be needed to improve both its potency and its specificity so that it might be further developed into a lead compound.

In collaboration with Dr. Jeff Aubé and his group, I have tested both commercially purchased analogs of SE-1 (Fig. 16) and SE-1 analogs synthesized by the Aubé group for inhibition of the AraC family activator VirF from *Shigella*. To do this, I used a whole cell based VirF-activated *virB-lacZ* fusion *in vivo*. Various parts of SE-1 were substituted to explore the structure–activity relationship (SAR) (Fig. 17). All the analogues were tested at various concentrations in the VirF reporter strain, as well as in a control strain. The control strain carries a LacI-repressed fusion and LacI-expressing plasmid. This fusion contains *lacZ* under the control of a synthetic promoter (P*hts*) with an induction level similar to that of the *virB* fusion [81]. Of all the analogues tested, we would screen for analogues that show greater inhibition (lower IC_{50}) value) than SE-1, which would suggest that these analogues have a much higher potency than SE-1. In addition, to consider an analogue to be a true positive hit, the analogues with greater inhibition also would need to have little or no inhibition against the control strain and does not inhibit bacterial cell growth, which would suggest that the hit analogue does not non-specifically target other cellular proteins and interfere with normal cell function.

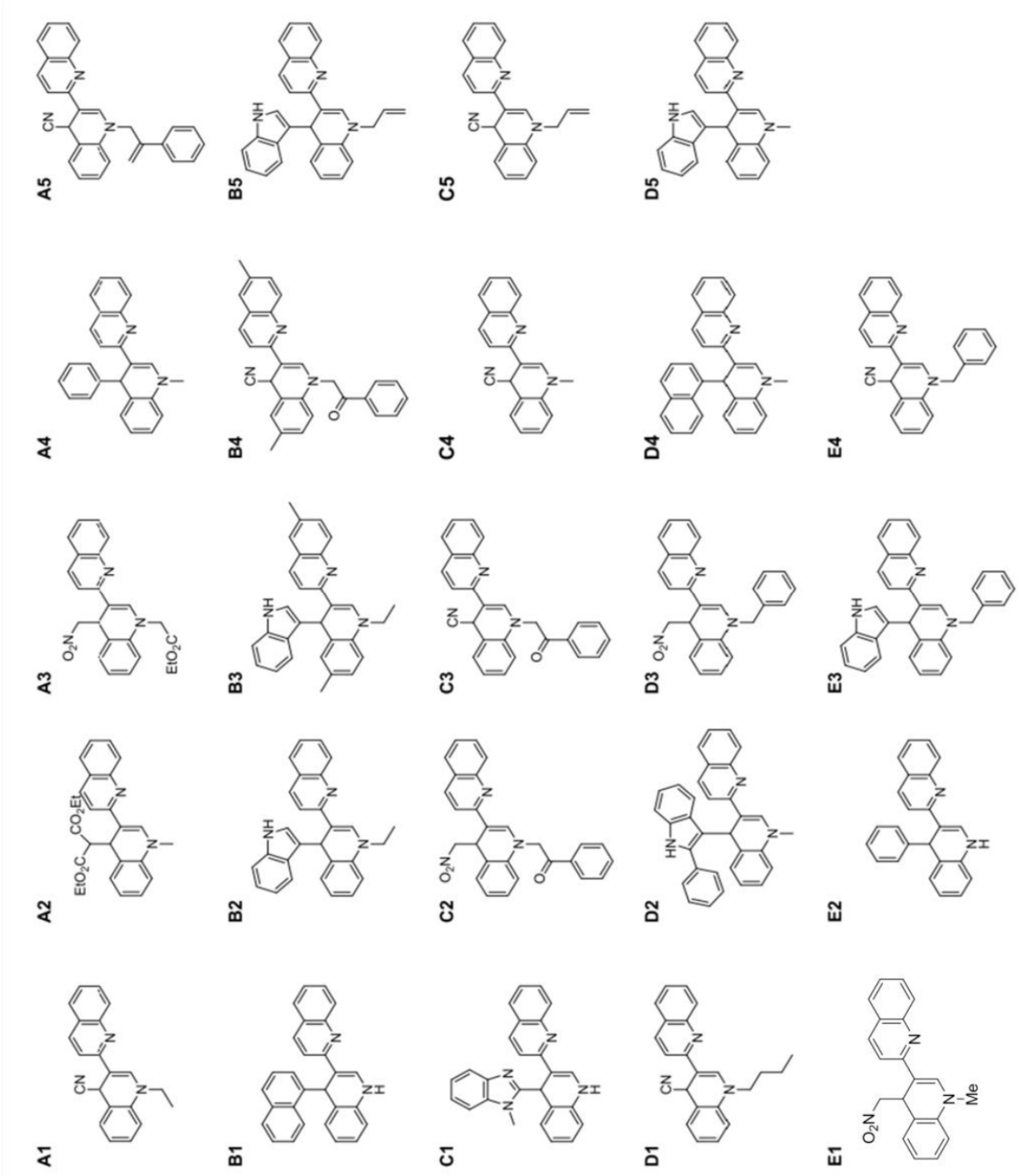


Fig. 16. Structures of all the commercially procured compounds are shown.

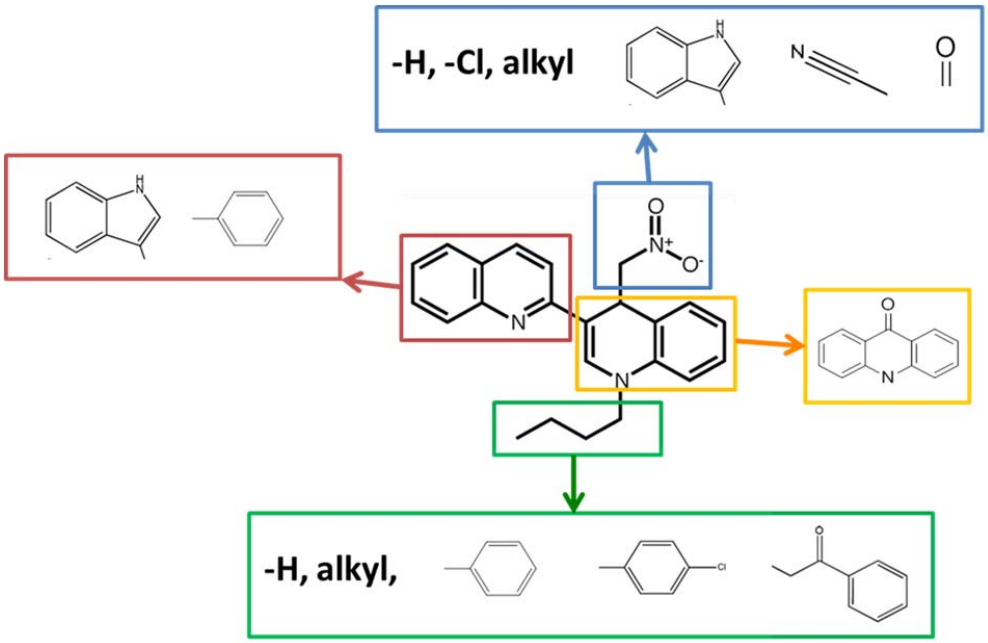


Fig. 17. Illustration showing that various parts of SE-1 were substituted with different chemical groups.

First, I have tested 24 commercially purchased analogues selected by the Aubé group to explore the initial SAR (Fig. 16), and found that analogues E1, and A2 inhibited VirF to approximately the same extent as SE-1, with IC_{50} values of 7 μM and 10 μM respectively; while analogues C4 and D3 exhibited less inhibition relative to SE-1, with IC_{50} values of 30 μM and 45 μM, respectively (Fig. 18 top). Other analogues with substitutions, such as an indole, a phenyl ring, a benzimidazole and a napthyl ring, exhibited no detectable inhibition of VirF in this assay (Fig. 18 bottom).

Next, seven 3,4-disubstitued dihydroquinolines, three quinolinium synthetic precursors were synthesized by the Aubé group, and tested in the whole cell based inhibition assay (entry 5-11, Table 9). It was found that replacing the *n*-butyl group with a 4-chlorobenzyl (entry 5, Table 9) caused a decreased inhibitory activity against VirF with IC_{50} of 80 μM. Analogues with the quinolone replaced by 5-methyl pyridine (entry 6, Table 9), 3-chlorophenyl (entry 7, Table 9)) also showed a significant loss of inhibition against VirF with $IC_{50} > 100$ μM. Analogues with the nitro-methyl group of the 1,4-dihydroquinoline replaced with phenyl (entry 8, Table 9), methyl (entry 11, Table 9) were found inactive, as well as analogues with the same substitution of a 1,2-dihydroquinolines (entries 9 &10, Table 9). Analogues with the quinoline core replaced were also synthesized and tested. The results showed that replacing the quinolone core with an indole (entries 12 and 13, Table 9), 9-acridone (entry 15 Table 9) and 4-quinolone (entry16, Table 9) resulted in no inhibitory activity against VirF.

Surprisingly, it was found by the Aubé group that SE-1 and many of the tested analogues (entries 5-11, Table 9) have exhibited noticeable instability since these analogues easily

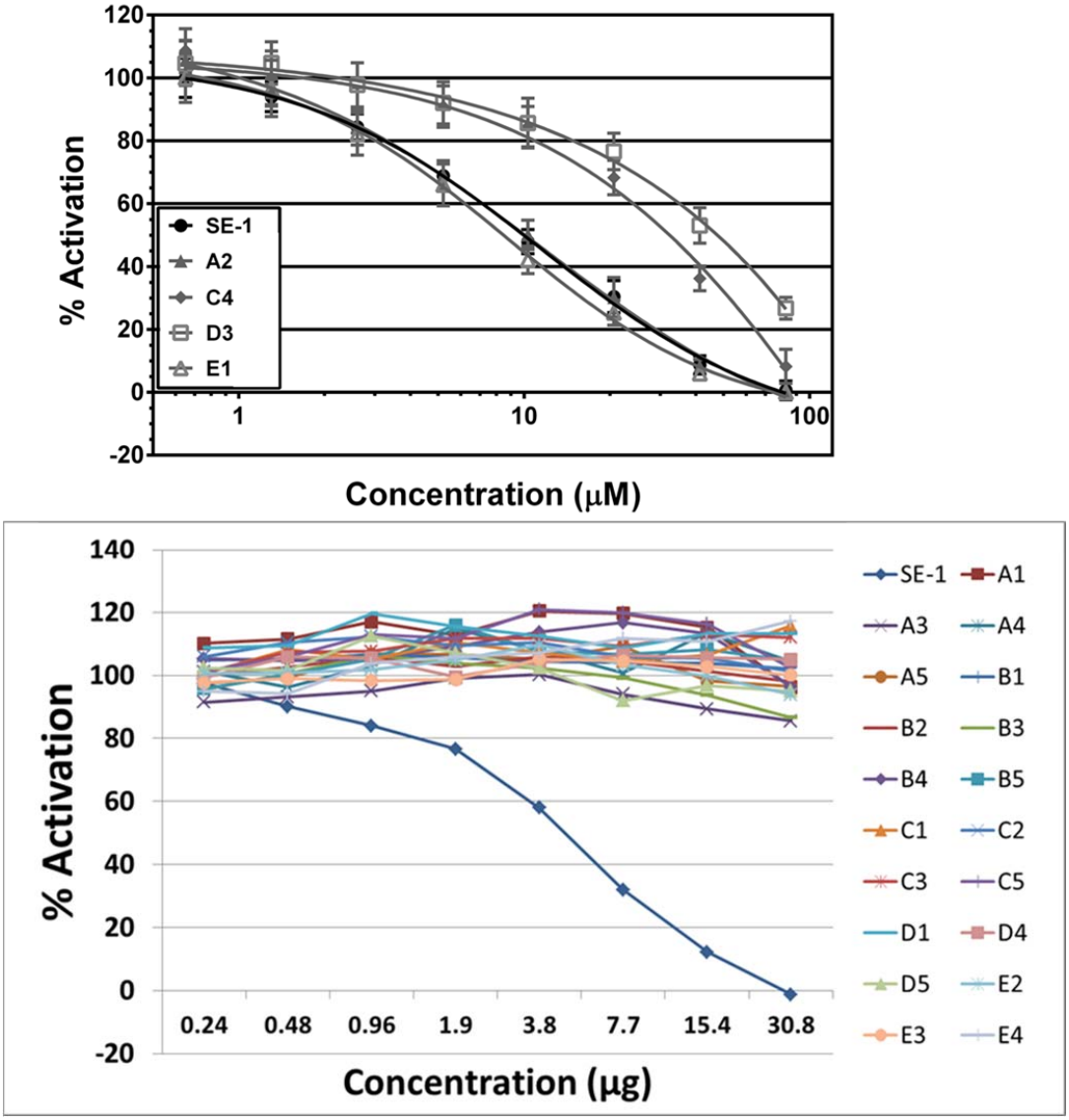


Fig. 18. *In vivo* inhibition of SE-1 and its analogues on β-galactosidase expression from the *virB-lacZ* fusion by VirF. Activity in the absence of SE-1was set to 100% for the reporter fusion. Results are the average of three independent experiments with two replicates each.

decomposed during their synthesis. In addition, it was observed that various dihydroquinoline analogues dissolved in DMSO at room temperature had reversed to the quinolinium precursor form after two days. Further analysis of the aqueous stability of the 3,4-disubsititued dihydroquinoline analogues was carried out. Analogues were dissolved in 100% DMSO first and then diluted to 10 μM in phosphate-buffered saline (PBS). Analogues samples were then analyzed every hour through an auto-sampler by RP HPLC/UV/HRMS for a time course of 48 hours at room temperature. It was found that SE-1 and many analogues converted completely to the corresponding quinoliumium salt upon immediate dissolving in the aqueous solution (Fig. 19). Given that there wasn't any detectable SE-1 in aqueous solution that had not converted to the quaternary salt, it was likely that the observed inhibitory effects of SE-1 and its analogues were caused by the quinolinium salt since the reporter assays were performed with SE-1 dissolved in aqueous buffer solution. To test this hypothesis, three quinolinium salts analogues (Table 9, entries 2–4) were tested, and the result showed that these quinolinium salts inhibited VirF activity to the same level as compared to their non-quinolinium salt form (Table 9, entries 2 to SE-1, entries 3 to 5, and entries 4 to E1). This result supported the hypothesis that the quinolinium salt is responsible for the observed inhibitory activity.

Given that many synthesized analogues converted completely to the corresponding quinoliumium salts form in aqueous solution, we want to be certain that the observed inhibitory activity by the quinolinium analogues was not caused by detergent effect. Some antibiotics, for instance, cetylpyridinium chloride (CPC) - a cationic quaternary ammonium salt, act as detergents and antiseptic agents. CPC binds to the negatively charged bacterial cell membrane.

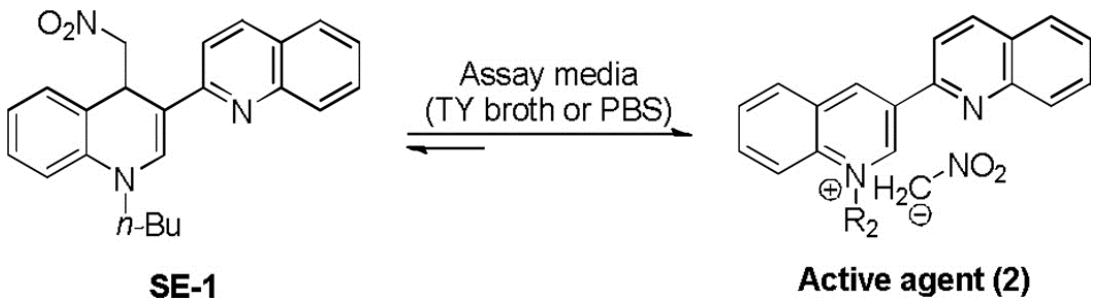


Fig. 19. *In situ* decomposition of SE-1 to quinolinium salt. From [130].

The nonpolar regions of CPC, which has similar characteristics to the cell membrane lipids, penetrate the cell membrane and thus alter the osmolality of the cell ultimately resulting cell death [131, 132]. Three analogues that behave similarly to CPC were synthesized and tested. The triethylammonium (entry 18, Table 9) analogue showed no inhibitory activity, the acridinium (entry 17, Table) and pyridinium (entry 18, Table 9) showed non-specific activity against both the VirF (8 μM and >50 μM, respectively) and control strain (15 μM and >50 μM, respectively), suggesting that the inhibitory activity of SE-1 and its analogues is not due to the detergent effect.

Further optimization of the quinolinium core was carried out by synthesizing and testing various N-substituted and 3-aryl analogues. It was found that analogues including *i*-propyl (entry 19, Table 9), *n*-propyl (entry 20, Table 9), *n*-hexyl (entry 21 Table 9) and pent-4-en-1-yl (entry 24, Table 9) all showed inhibitory activity similar to SE-1 with IC_{50} values of 12 μM, 16 μM, 11 μM and 11 μM respectively; the exception was the methyl substitution (entry 4, Table 9) which exhibited a decreased inhibitory activity with a IC_{50} value of 90 μM. In addition, analogue with 3-quinoline substituent replaced by naphthalene group (entry 23, Table 9) also showed similar inhibitory activity to SE-1against VirF (9 μM of IC_{50}). However, analogue with *n*-nonyl (entry 22, Table 9) showed increased inhibition with an IC_{50} value of 3 μM against VirF and an IC_{50} value of 25 μM against the control strain, which suggested that this analogue may have non-specifically bound to the cellular proteins and interfered with normal cell functions. Therefore, *E. coli* cell growth study was carried out to further analyze the cell toxicity of this n-nonyl (entry 22, Table 9) analogue. It was found that n-nonyl analogue inhibited cell growths

Entry	Structure	VirF IC_{50} (μM)	Control IC_{50} (μM)
SE-1/1	-	11	>100
2		11	>100
3		26	>100
4		10	>100
5		80	>100
6		>100	>100
7		>100	>100
8		No activity	No activity
9		>100	>100
10		No activity	No activity
11		>100	>100
12		No activity	No activity
13		No activity	No activity
14		8	15
15		No activity	No activity
16		No activity	No activity
17		>50	>50
18		No activity	No activity
19		12	>100
20		16	>100
21		11	>100
22		3	25
23		9	>100
24		11	>100

Table 9. IC_{50} of synthesized analogues tested in the strain with VirF-activated *lacZ* fusion and a control strain with a synthetic promoter driving the expression of *lacZ.* Results are the average of three independent experiments with two replicates each.

of both the VirF and control strains as compared to other analogues (both synthetized and purchased), suggesting the cell toxicity of this analogue.

Given that the quinolinium salts are electrophilic in nature, and can accept an electron pair in order to bond to a nucleophile [133]. Once the quinolinium salts bonded with a nucleophile species forming new compounds, their chemical property may be altered. This could result in losing selectivity towards the target protein (for instance, VirF or RhaS) since the new compound may not be able to bind to the target protein or function differently. However, recent research shows that Gram negative antibiotics tend to be positively charged, and the charge is required for uptake of many known Gram negative antibiotics, for example the aminoglycosides including streptomycin and gentamicin [134, 135]. The positive charge helps this class of antibiotics entering the Gram negative bacteria cell, and primary inhibiting protein synthesis [134, 135]. Since all the analogues were initially tested in whole cell based assays, it is likely that some analogues without charges may not show inhibition in this assay due to lack of uptake into the bacterial cells. Therefore, I have tested uncharged analogues synthesized by Aubé group using a cell-free assay – an *in vitro* DNA-binding EMSA assay using purified RhaR protein (Fig. 20). This assay directly assays the inhibitory effect of each analogue on protein, and does not require the uptake of the analogue. I quantified each gel using ImageJ [136], to determine the percentage decrease in DNA binding by RhaR in the presence of 82.5 mM final concentration of each analog. Using this assay, I identified thirteen analogues showing at least 20% inhibition of DNA-binding by RhaR (Table 10). Some analogues showed more inhibition *in vitro* than *in vivo*, including analogues A1, C5, E4, A17, and A23, which suggested that uptake of uncharged analogues may be a reason why some analogues showed no or little inhibition in

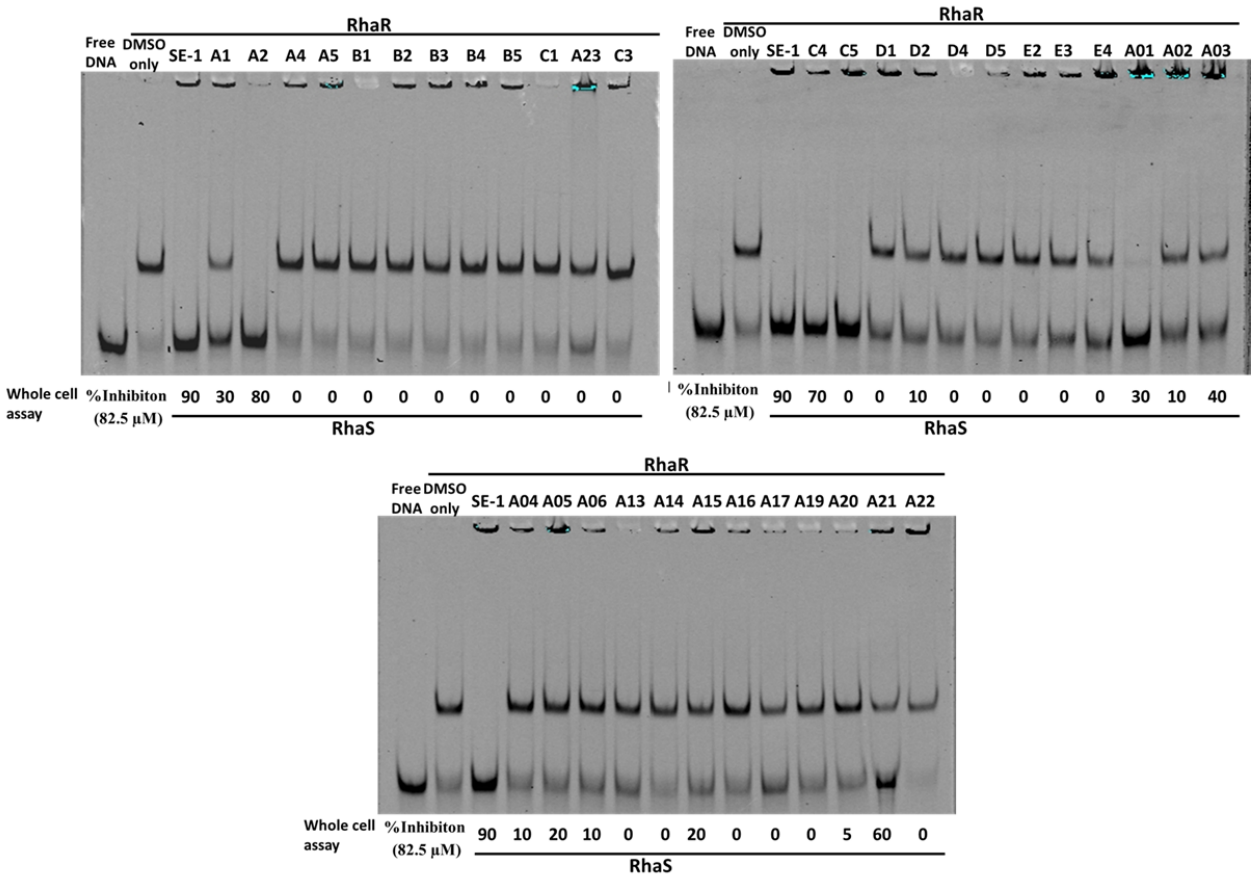


Fig. 20. SE-1 and its analogues inhibited DNA-binding by RhaR *in vitro*. A set of representative EMSA gel images were shown. Analogue concentrations were in large excess as compare to protein concentration, and between 1.1 to 2.0mM (except for E4: 0.65 mM).

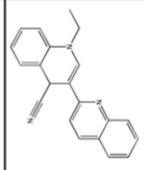
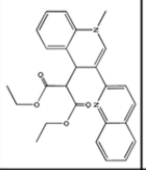
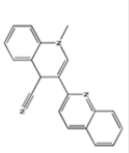
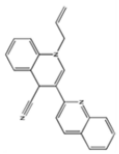
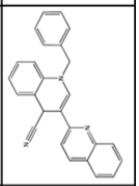
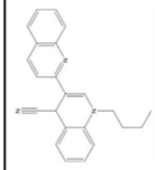
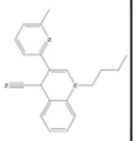
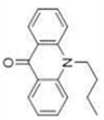
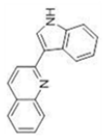
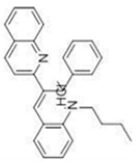
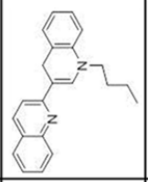
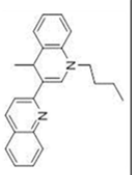
Entry	Structure	RhaR EMSA %Inhibition	RhaS Whole cell assay %Inhibition
A1		80	30
A2		100	80
C4		100	70
C5		100	0
E4		20	0
A01		100	30
A02		40	10
A03		40	40
A13		30	0
A15		20	20
A17		35	0
A21		50	60
A23		30	0

Table 10 IC_{50} of analogues tested in the *in vitro* DNA-binding by RhaR, and *in vivo* whole cell assay with RhaS. Results are the average of three independent experiments with two replicates each. Analogues concentrations were between 1.1 to 2.0mM (except for E4: 0.65 mM). Results were the average of three independent experiments with two replicates each.

the *in vivo* whole cell-based assays. Analogue A21 was very similar to SE-1 in structure, but was uncharged and showed 2-fold less potent inhibition as compared to SE-1, suggesting that the positive charge may also contribute to the inhibition.

Chapter 6

Purification of RhaR and Identification of RhaR annotation Error

RhaR is a member of the AraC family of bacterial transcription regulators, and activates transcription of the *rhaSR* operon of the *E. coli* L-rhamnose regulon in the presence of L-rhamnose [10, 46, 53]. In order to better understand the properties of RhaR, and the mechanism of transcription activation by RhaR, we need to purify RhaR protein to perform biochemical assays. Previously, untagged RhaR protein has been purified with L-rhamnose using DNA affinity chromatography, and the protein purified was active [53]. However, the yield of RhaR protein was quite low; only 0.1 mg protein was purified from 6.6 L of *E. coli* [53]. Our lab has constructed the chitin-binding domain-intein fusion protein, and the yield of RhaR protein was extremely low (10 μg protein was purified from 1 L of *E. coli*) despite that the protein was active. We later constructed GB1 fusion protein, and the His_6-$GB1^{basic}$-RhaR was soluble and active. The yield of RhaR protein using this construct was relatively high (at least 1mg per 1 L *E. coli*). However, since the GB1 tag was at the N-terminus of the protein, it might interfere with the conformational change of the arm region in response to rhamnose and thus the fusion protein may not be as active as it supposed to be. Here, I initially purified RhaR by using the chitin binding domain (CBD)-intein fusion. In the process of working with that fusion protein, I discovered that there was an error in the annotation of *rhaR*. Once the correct *rhaR* open reading frame was overexpressed, there was a dramatic increase in the overexpression and solubility of active RhaR protein. For simplicity, we will refer to the native gene and protein as *rhaR* and RhaR, and the previously annotated, non-native gene and protein as +30*rhaR* and +30RhaR.

Purification of RhaR using CBD-intein fusion. RhaR protein was first expressed as the CBD-intein-RhaR fusion protein, and loaded onto a chitin column for purification. The chitin resin with CBD-intein-RhaR bound was mixed with buffer containing 100mM DTT to cleave the untagged RhaR (now shown to be +30RhaR) protein off the column, and untagged +30RhaR

protein was subsequently eluted (Fig. 21A). Using this method, the +30RhaR protein I obtained was active based on its ability to bind to a DNA fragment carrying its specific DNA recognition sequence when assayed by EMSA in the presence of L-rhamnose (Fig. 21C). However, the protein wasn't pure as shown on the 12% SDS-PAGE that showed two bands at approximately the size of RhaR (Fig. 21B). In order to determine which band was RhaR, western blots were performed using custom anti-RhaR antibody and a control sample of RhaR. Surprisingly, we found that both of the bands reacted with the anti-RhaR antibody, (Fig. 21D), suggesting there might be two variants of RhaR in the purified RhaR protein (+30RhaR and RhaR). A major difference between RhaR and RhaS, and indeed between RhaR and most of the other AraC family proteins, is that RhaR contains a 33 amino acid extension at the N-terminal end. The calculated molecular mass of the extension was around 4 kDa, and the size difference between the two RhaR bands shown on the gel approximately matched the size of the extension. In addition, earlier in our lab, we constructed and assayed various N-terminal deletions within this extension region, and found that RhaR Δ29 and Δ34 (deleting most or all of the extension region) migrated to the same position on SDS-PAGE as wild-type RhaR; only the RhaR Δ40 ran at a smaller size as compared to the wild-type RhaR in western blot (Fig. 22). This result, combined with the western blot of the purified RhaR protein suggested that the wild-type RhaR may not include the extension. Therefore, we hypothesized that the bottom band of RhaR might be a natural cleavage product of RhaR, likely removing residues from the N-terminal end. The purity from this purification was quite poor, and the yield was quite low.

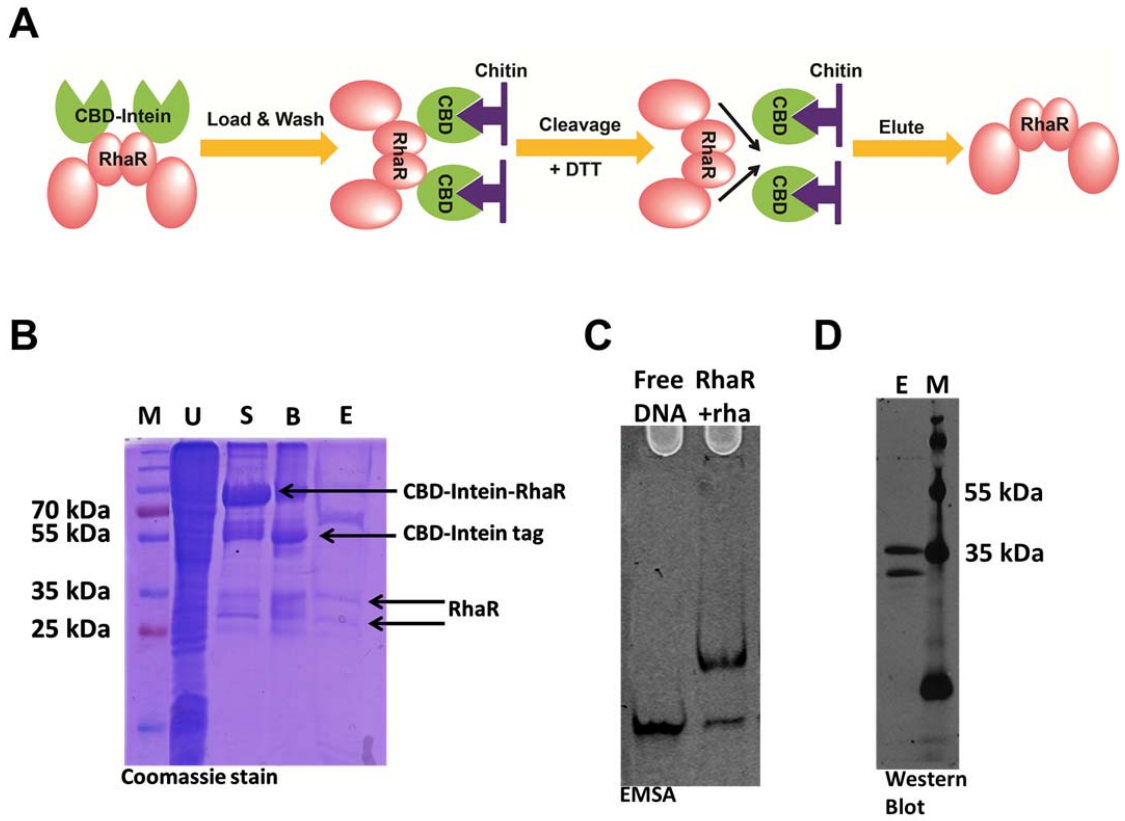


Fig. 21. **(A)** Schematic showing the procedures of RhaR CBD-intein fusion purification. **(B)** 12% SDS-PAGE gel showing purified RhaR protein with two sizes. M, molecular mass marker; U, uninduced sample; S, supernatant; B, beads after DTT cleavage; E, concentrated protein elution. **(C)** The EMSA gel showing the DNA-binding by purified RhaR with L-rhamnose (+rha). **(D)** Western Blot of purified RhaR protein. E, concentrated RhaR protein elution; M, molecular mass marker.

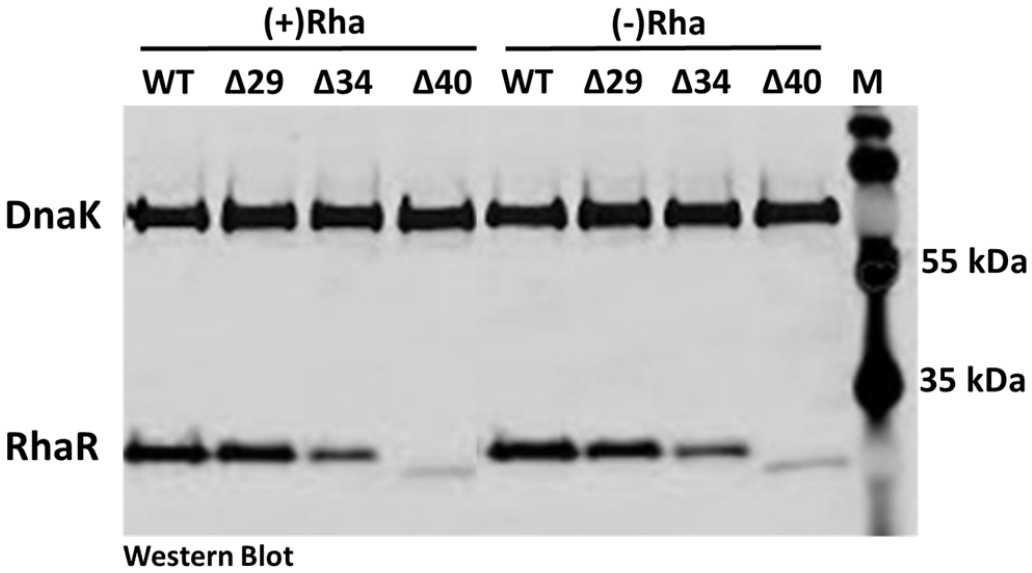


Fig. 22. Western blot showing the protein levels and relative sizes of RhaR N-terminal deletion variants. The culture samples were grown in the presence and absence of L-rhamnose [(+)rha and (-)rha, respectively]. Equivalent cell densities (based on OD_{600}) of the culture samples were loaded in all of the lanes. All RhaR N-terminal deletion variants were expressed from pHG165. Molecular mass marker (M) is in the last lane.

Purification of RhaR using His_6 tagged fusion. Because of the complications with the CBD-intein-RhaR fusion, I attempted to purify RhaR using a different construct which was the C-terminal His_6-tagged RhaR fusion. Since we hypothesized that there might be a cleavage near the N-terminal end of RhaR, we expected that we might be able to enrich for the smaller sized RhaR by using a C-terminally tagged fusion protein. Using this construct, I was able to purify the RhaR protein (now shown to be +30RhaR), however, two different sized RhaR proteins (+30RhaR and RhaR) were also purified from this construct, as observed from the 12% SDS-PAGE gel and western blot (Fig. 23A & B). The purified protein was active, based on the evidence that it bound to the DNA in an EMSA, as shown in Fig. 23C. In order to determine which form of the RhaR protein was active, I performed the DNA binding assay in EMSA by using protein from elution fraction 28 (F28) which is enriched for the larger sized RhaR (+30RhaR), and fraction 33 (F33) which is enriched for the smaller sized RhaR (RhaR) (Fig. 23D). There was much more active RhaR protein in fraction F33 than in F28, suggesting that the smaller sized RhaR (RhaR) was the active form. Additional support for this hypothesis that the smaller sized RhaR (RhaR) is the active form came from the first purified RhaR (RhaR) using DNA affinity purification as reported by Tobin and Schleif [53]. This purification was expected to purify only active protein, since it required that the protein bind to the DNA that had been linked to the resin. The size of the active RhaR (RhaR) purified by Tobin and Schleif [53] was roughly the same size as the smaller sized RhaR (RhaR) (Fig. 24). These results support the hypothesis that the active RhaR (RhaR) is the smaller sized band on SDS-PAGE gels.

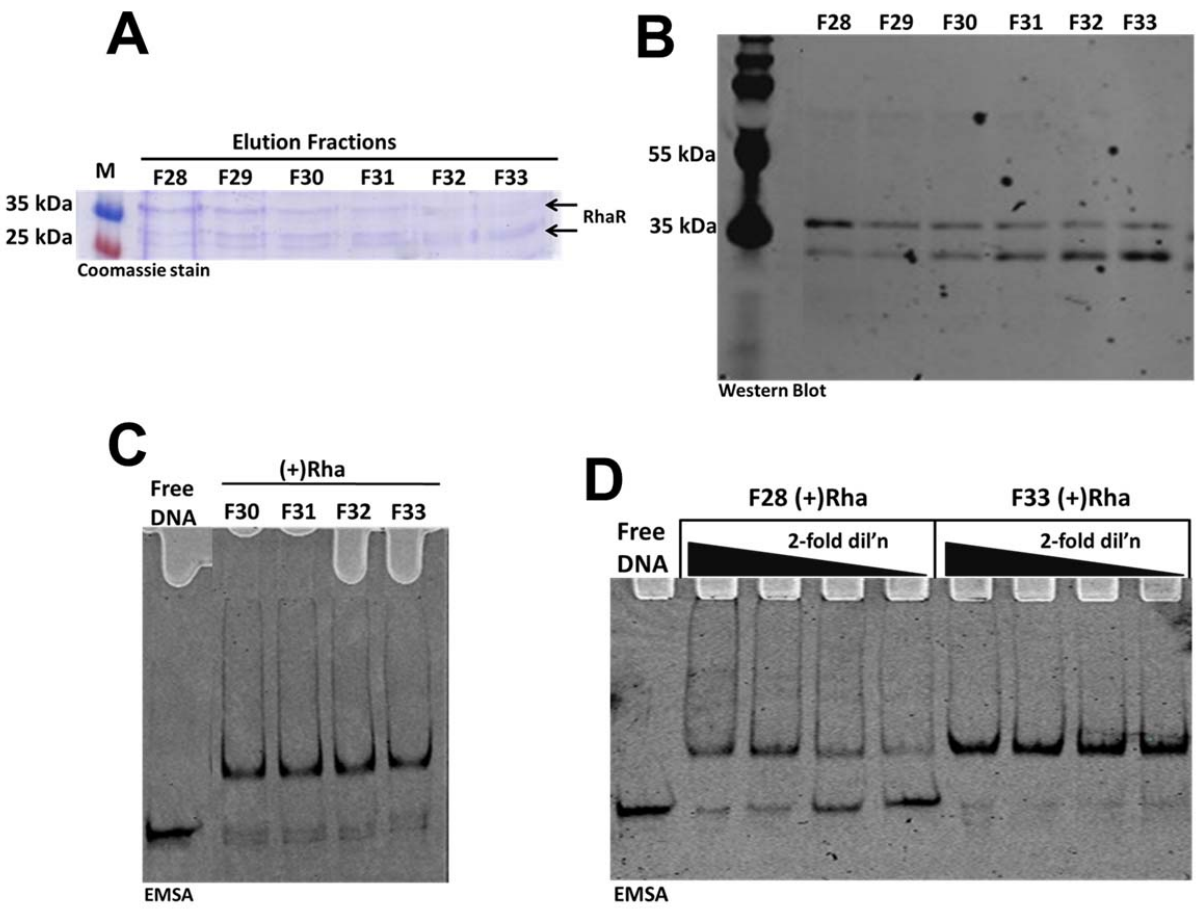


Fig. 23. **(A)** 12%SDS-PAGE gel showing purified RhaR-His_6 protein with two different sizes. Molecular mass marker (M) is in lane 1; lane 2 to 6 are elution fractions from 28 to 33. Two bands that both react with anti-RhaR antibodies (part B) are indicated by arrows. **(B)** Western blot showing the two bands in the partially purified RhaR-His_6 that react with anti-RhaR antibodies. Molecular mass marker (M) is in lane 1; lane 2 to 6 are elution fractions from 28 to 33. **(C)** DNA-binding by purified RhaR-His_6 protein fractions from 28 to 33 with L-rhamnose [(+)Rha] in EMSA. **(D)** EMSA showing DNA-binding activity of purified RhaR-His_6 fraction 28 (F28) enriched for larger sized RhaR, fraction 33 (F33) enriched for smaller sized RhaR and their dilutions in the presence of L-rhamnose [(+) rha]. Black triangles represent decreasing concentrations of proteins, with serial two-fold dilutions. Free DNA (F) is in first lane.

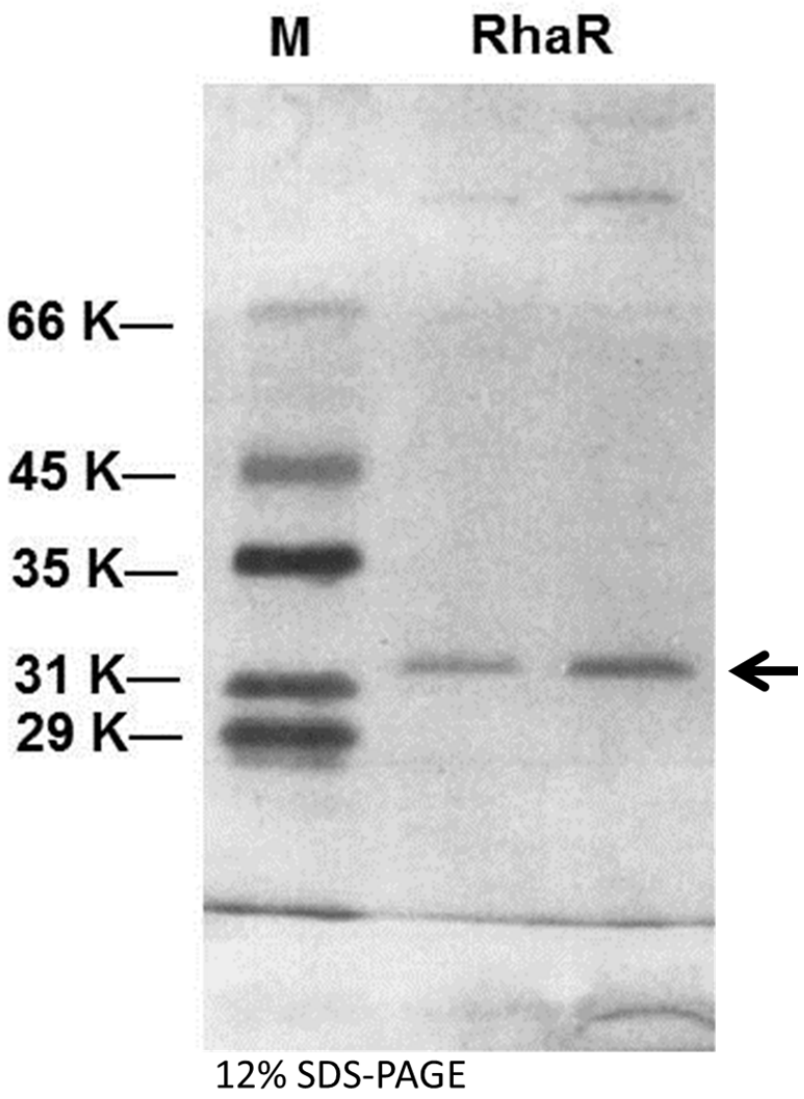


Fig. 24. 12% SDA-PAGE gel of purified RhaR. The arrow indicates the location of RhaR protein. M, molecular mass marker. From Tobin & Schleif (1990) [53]

Protein peptide mapping and N-terminal protein sequencing. In order to test the hypothesis that the smaller and larger sized RhaR proteins (RhaR and +30RhaR) differed at their the N-terminal ends, samples of the protein purified from the His_6 RhaR fusion (RhaR and +30RhaR) were sent to the Mass Spectrometry and Analytical Proteomics Laboratory at the University of Kansas, where they performed protein in-gel digestion and analyzed the resulting peptides by mass spectrometer. Several of the same RhaR peptides were not detected from both the longer and shorter bands, suggesting that these did not correspond to the difference between the proteins. The only difference in the peptide maps was near the N-terminus, where the larger band included peptides corresponding to the previously annotated RhaR sequence minus only the N-terminal methionine (predicted 312 residues) (+30RhaR), while the smaller protein had no peptide coverage for the first 36 residues of the previously annotated RhaR sequence (RhaR, Fig. 25). This result supported the hypothesis that the difference between the two proteins was at the N-terminus. To determine the precise N-terminal end of the smaller sized RhaR, it was sent for N-terminal sequencing by Edman degradation to the protein facility at Iowa State University. The result indicated that the smaller protein was missing the first 31 amino acids relative to the prior annotation, with the sequence Ala-His-Gln-Leu-Lys at the N-terminus (predicted 282 residues). This allowed us to identify the N-terminus of the RhaR that our assays indicated was the active form of RhaR.

Identification of RhaR annotation error. Examination of the sequence of the *rhaR* open reading frame showed that the codon prior to that encoding the N-terminal alanine

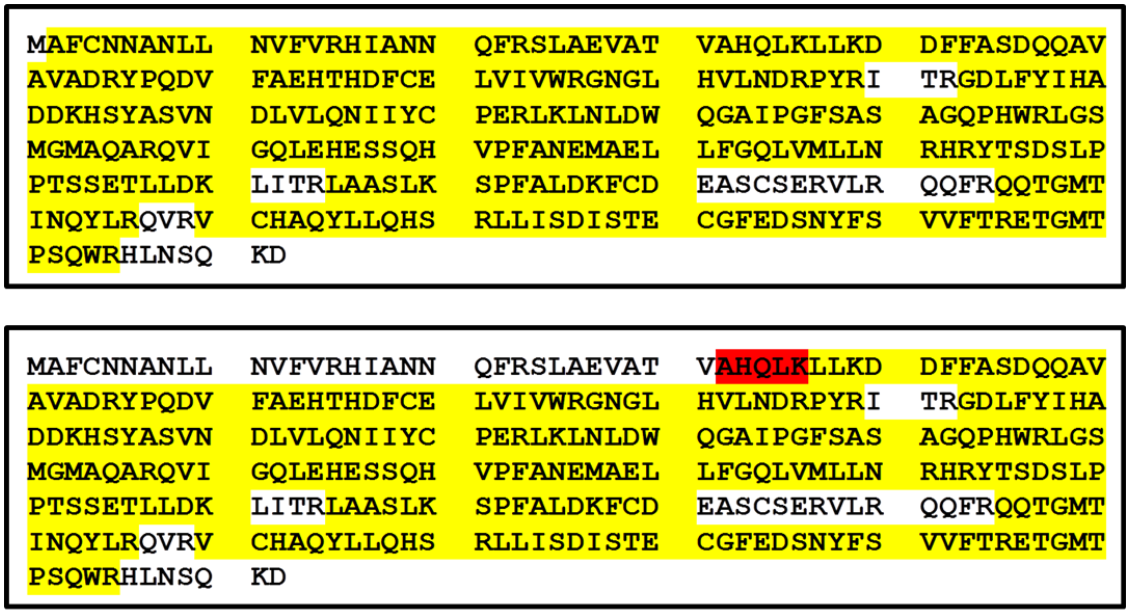


Fig. 25. Peptide coverage of two different RhaR protein products determined by LC-MS/MS. Top, peptide sequence coverage of the larger sized RhaR protein. Bottom, peptide sequence coverage of the smaller sized RhaR. Peptides detected by LC-MS/MS are highlighted in yellow. The N-terminal sequence of the smaller protein, determined by Edman degradation, is shown in red.

was GTG, which is used as a start codon in 14% of *E. coli* genes [137]. Additionally, a potential Shine-Dalgarno (SD) sequence (GAGG) was identified 8 bp upstream of the putative GTG start codon (Fig. 26). Taken together with the finding that the smaller, 282-residue RhaR protein was more active in DNA binding, these results suggested that the true *rhaR* translational start codon might be GTG, and that the first methionine was removed after translation in vivo. This was consistent with the sequence obtained by Edman degradation, and would produce a protein with 281 residues and a molecular mass of 32 kDa. These results also suggested that the 282-residue RhaR protein may not be a product of cleavage, but rather was the product of a start codon that was different from the original annotation.

Additional support for the hypothesis that RhaR (282-residue) represented the native RhaR protein came from analysis of plasmids designed for *in vivo* RhaR transcription activation assays (rather than high-level overexpression). In contrast to pET24b (and other "pET" vectors), which provide an optimally positioned, strong SD signal, plasmids for in *vivo* RhaR expression (on the pHG165 vector) relied on the SD sequence provided by the *rhaR* region. Despite the presence of the upstream (originally annotated) ATG start codon and associated putative SD sequence, we found that pHG165-based +30*rhaR* clones expressed a protein the size of native RhaR (~282 residues), consistent with translation beginning at the downstream GTG codon (Fig. 21D and Fig. 22). Thus, it appears that the putative upstream SD sequence (positioned near the +30*rhaR* ATG start codon) directed little or no translation, as determined by little or no detection of +30RhaR protein by Western blot. However, upon correctly positioning the strong pET24b SD upstream of the +30*rhaR* ATG start codon "forced" translation to initiate predominantly

Fig. 26. Portions of the *rhaR* open reading frames with upstream SD sequences. The previously annotated SD sequence of RhaR (312 residues) is underlined in blue, and the corresponding protein sequence is also in blue. The upstream SD sequence for the true RhaR (282 residues) is underlined in red, and related protein sequence is also shown in red.

at the non-native, upstream, in-frame, ATG start codon. RhaR expressed from pHG165 activated transcription to a level that was six-fold higher than the activation by RhaR expressed from the native chromosomal locus, consistent with the modest overexpression of RhaR from pHG165, and suggesting that RhaR expressed from pHG165 was fully functional [10, 57]. Importantly, these findings indicate that RhaR expressed from pHG165 was the native RhaR sequence, regardless of whether +30*rhaR* or *rhaR* was cloned. Thus, *in vivo* experiments reported by our lab and the Schleif lab prior to this publication erred only in the numbering of RhaR residue positions, with an extra 30 residues at the N-terminus of RhaR relative to the native protein sequence.

As an additional test of the hypothesis that the downstream GTG codon served as the native RhaR start codon, I have constructed a pHG165-based clone to express *rhaR* from the GTG start codon and corresponding SD sequence (pSE312). A construct with one additional nucleotide introduced immediately after the ATG start codon was also made. In this construct, the expression of +30*rhaR*, if any, from the upstream ATG will be out of frame (pSE313). However, if translation starts at the GTG codon, the amount of active RhaR protein expressed, and transcription activation by RhaR will not be effected. A clone with the previously annotated start codon and its associated SD sequence (pSE302) was also included in this experiment. The results showed that *in vivo* RhaR transcriptional activation activities, protein levels, and protein size were indistinguishable from the three plasmids (pSE302, pSE312, pSE313) (Fig.27). These results support the hypothesis that RhaR translation initiates at the GTG codon, not the ATG codon, and that the previously annotated RhaR protein sequence had 30 extra amino acids at its N-terminal end relative to the native 282 residue RhaR protein sequence.

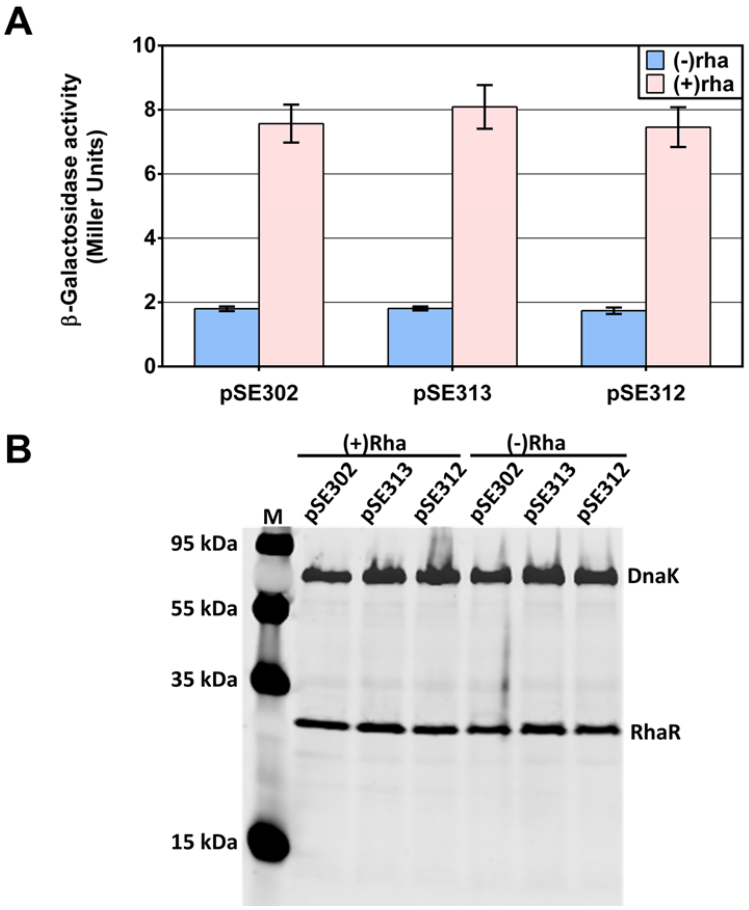


Fig. 27. (A) β-Galactosidase activity was assayed from a single-copy Φ(*rhaS-lacZ*)Δ85 fusion in a strain with Δ(*rhaSR*)::*kan recA*::*cat* and RhaR proteins were expressed from a plasmid pHG165. Cultures were grown with or without L-rhamnose. pSE302 contained the previously annotated ATG start codon and associated SD sequence; pSE312 contained the GTG start codon and corresponding SD; pSE313 contained an extra nucleotide after the upstream ATG start codon. (B) Western blots showing the protein levels and sizes of RhaR protein expressed from each of the three plasmids. The culture samples in all of the lanes were loaded at equivalent cell densities (based on OD_{600}). DnaK was used as a loading control. Molecular mass marker is in the first lane.

Purification of His_6 tagged RhaR having the newly identified start codon. I constructed a new RhaR overexpression clone in pET24b (pSE308) to express the true RhaR protein sequence. This plasmid expressed a much larger amount of soluble RhaR protein, and only a single band corresponding to RhaR was detected by SDS-PAGE. After purification of the protein by Ni^{+2}-affinity and size exclusion chromatography, we obtained roughly 10 mg of soluble RhaR protein per liter of *E.coli* cells grown in the absence of rhamnose, and the RhaR protein was at least 95% pure. No peak at the expected mass of RhaR monomer was detected, suggesting that the protein was in the form of a dimer. The protein was also active, as measured by its binding to the RhaR binding site DNA in an EMSA (Fig. 28).

RhaR was slightly more stable and soluble when the overexpressing cells were grown and the protein was purified in the presence of rhamnose. However, RhaR protein from cells grown in the absence of rhamnose and also purified in the absence of rhamnose was sufficiently stable and soluble for the purposes of this study, and allowed direct comparison of activity in the absence and presence of rhamnose.

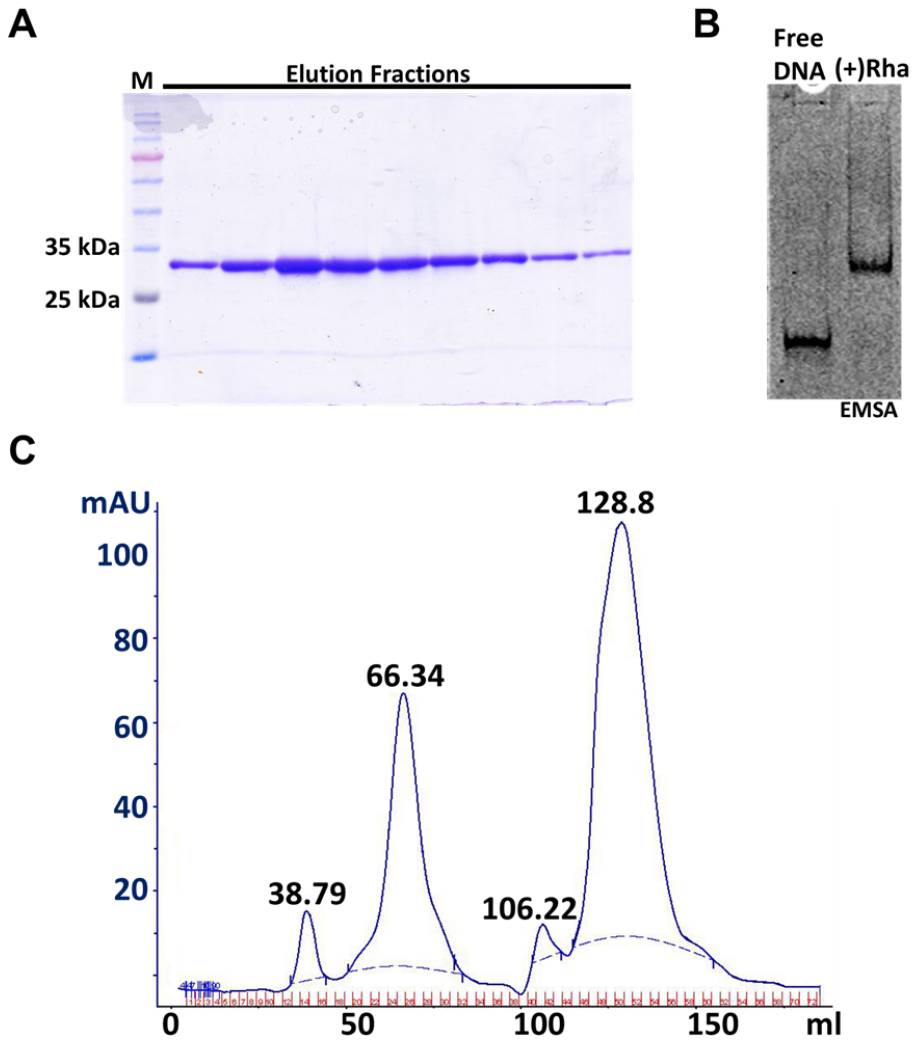


Fig. 28. **(A)** 12%SDS-PAGE gel showing purified RhaR-His6 protein eluted from size exclusion chromatography. Molecular mass marker (M) is in the first line. **(B)** DNA-binding by purified RhaR-His_6 with L-rhamnose [(+)Rha] in EMSA. **(C)** Size exclusion chromatography of RhaR-His_6. Y-axis: UV absorbance at 280 nm and X-axis shows elution volume. The peak at 38.79 ml retention represents aggregation. The next peak at 66.34 ml retention eluted at the size of ~64 kDa, peaks at 106.22 ml and 128.8 ml retention eluted at the size of ~12 kDa and 0.12 kDa.

Chapter 7

Structure of the Regulatory Domain of RhaR with and without its effector rhamnose bound

Despite the large size and ubiquity of the AraC family, only a relatively few members have been well characterized, largely due to the tendency of family members to exhibit low protein solubility and difficulties in purification and crystallization. Thus, available high resolution structures of AraC family proteins are limited in number. Among the subset of AraC-family proteins that share sequence similarity with the RhaR regulatory domain, available structures include: the NTD of AraC, the transcriptional activator of the *E. coli* arabinose regulon [33, 34]; the NTD of ExsA, a transcriptional activator of type three secretion system genes in *P. aeruginosa* [58, 129]; and the NTD from the structure of full-length ToxT, which regulates the expression of cholera toxin and other virulence factors of *V. cholerae* [22, 23, 38]. In order to identify the mechanism allosteric signaling by RhaR, and to provide structural information, I have purified RhaR-NTD-His_6 protein, set up crystallization screens, optimized crystallization condition, and solved the structure of RhaR-NTD both with and without bound rhamnose.

Purification and crystallization of RhaR-NTD-His_6. The protein was overexpressed as a C-terminal His_6 tagged fusion protein from pET24b vector in *E.coli* BL21 (DE3) cells grown in the absence of rhamnose for initial crystallization screening. A two-step protocol was used to purify RhaR-NTD-His_6 protein. First, the protein was purified via Ni^{2+}-affinity chromatography, and the elution fractions showed that the protein was at least 90% pure. Elution fractions containing RhaR-NTD-His_6 were pooled and then further purified through size exclusion chromatography (SEC). After the SEC, the RhaR-NTD-His_6 protein was at least 95% pure (Fig 29). The protein elution fractions from the SEC were combined and concentrated to ~6 mg/ml for crystallization screens.

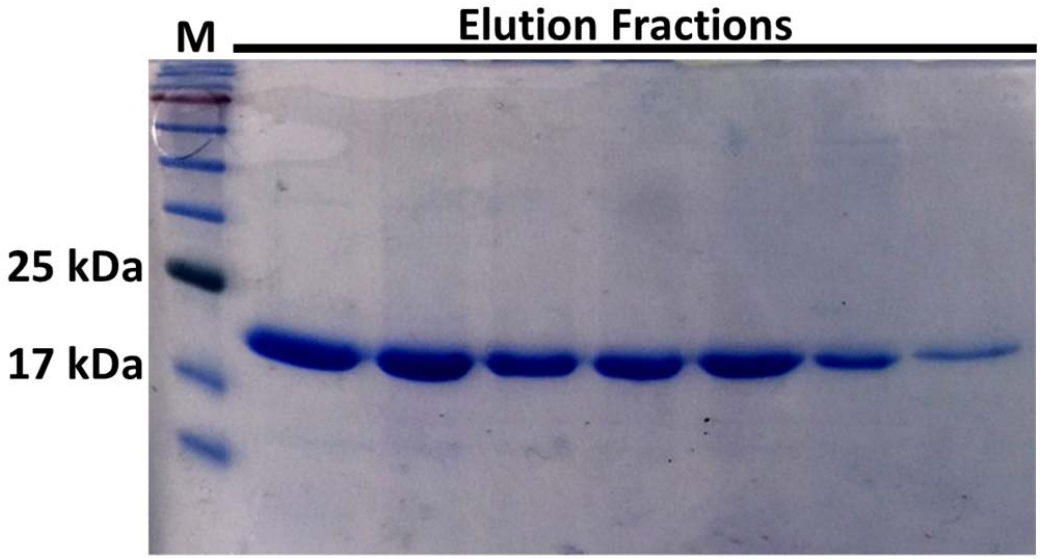


Fig. 29. 12% SDS-PAGE gel stained with coomassie blue showing purified RhaR-NTD-His_6 protein eluted from size exclusion chromatography. Molecular mass marker (M) is in the first lane.

Crystallization trials were carried out using commercially available crystallization kits Wizard I and II (Rigaku, Bainbridge Island, WA) in 96 well MarC 2 Well crystallization plates (Hampton research, Aliso Viejo, CA). For each condition, two sitting drops were set up; one drop with 0.2% L-rhamnose added to the protein, and the other with no added L-rhamnose. Among all 96 conditions tested, two conditions from the Wizard II kit produced RhaR-NTD-His_6 crystals without L- rhamnose. No RhaR-NTD-His_6 crystals were found in the presence of L-rhamnose. Of the two conditions that resulted in formation of crystals, condition 28 contained 20% (w/v) polyethylene glycol (PEG)-8000, 100 mM MES pH 6.0, and 200 mM $Ca(OAc)_2$, and the other, condition 48, contained 1 M K/Na tartrate and 100 mM MES pH6.0. Crystals from both conditions were small; therefore crystallization condition optimization was required to obtain larger crystals.

For each of the two conditions that resulted in crystal formation, condition optimization was performed. Various length oligomers of PEG (PEG-8000, PEG3350), several concentrations of PEG (from 6% to 20%), $Ca(OAc)_2$ (50 mM, 100mM, and 200 mM) and pH of MES (pH 5.5, 6.0, 6.5, 7.0) were tested in hanging drops using 24-well VDX plate (Hampton research, Aliso Viejo, CA). Wells with 6%-10% PEG-3350, 100 mM MES pH 6.0, 200 mM KCl, 50 mM $Ca(OAc)_2$, produced the largest crystals. Surprisingly, solutions in which the concentration of K/Na tartrate and the pH of MES were varied did not yield any crystals, and even repeating the original condition did not produce any crystal. It is unclear why obtaining crystals from condition 48 cannot be repeated. The large crystals were then picked, transferred to cryo-protectant, flash frozen with liquid nitrogen, and then sent to the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL) for X-ray diffraction. Crystal

diffraction data was collected and analyzed. One of the crystal diffracted to 2 Å, however, due to the phase problem, structure solving was not successful.

In order to solve the structure of RhaR-NTD-His_6 and the phase problem, seleno-methionine incorporated protein was need. Two protocols were used to overexpress seleno-methionine labelled RhaR-NTD-His_6. One method utilized the feedback inhibition of methionine biosynthesis by adding an amino acid mix including lysine, threonine, phenylalanine, leucine, isoleucine, and valine prior to induction. The advantage of this method is that the overexpression can be carried out in the standard BL21(DE3) strain. In order to confirm seleno-methionine had been successfully incorporated into purified protein from cells grown using this method, the putative seleno-methionine-labeled purified protein sample as well as the native (unlabeled) protein were sent to the Mass Spectrometry and Analytical Proteomics Laboratory at the University of Kansas to measure the molecular mass of each sample of protein (Fig. 30). The analysis showed that the molecular mass of the putative seleno-methionine-labeled RhaR-NTD-His_6 protein sample was 21,269 Da, which matched with the molecular mass of the native protein. This result indicated that the protein was not labeled with seleno-methionine.

The second method of seleno-methionine incorporation used a protocol from New England Biolabs (Ipswich, MA), with modifications. The RhaR-NTD protein was overexpressed in *E.coli* B834 cells (Novogen), a strain that is a methionine auxotroph. The cells were grown in minimal medium with regular L-methionine until the cell density reached 0.6, then pelleted and re-suspended in a minimal medium without any methionine. The cell culture was incubated for an additional 3 hours to deplete any residual methionine, and then seleno-methionine was added to the culture along with the inducer IPTG. Using this method, I was able to obtain RhaR-NTD

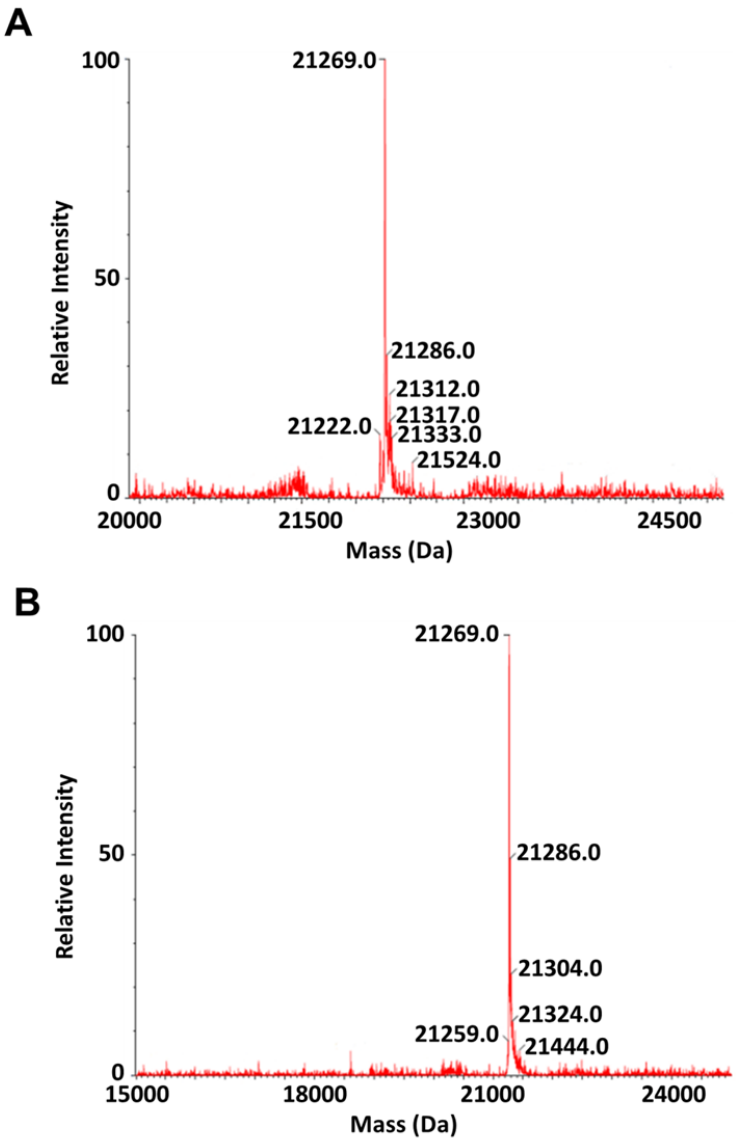


Fig. 30. The molecular mass profile of RhaR-NTD protein (A) RhaR-NTD-His_6 protein was overexpressed using feedback inhibition of methionine biosynthesis in an effort to incorporate seleno-methionine. The most intense peak was at 21,269 Da, a mass that was identical to the unlabeled protein. **(B)** Native RhaR-NTD-His_6 without any seleno-methionine labeling. The most intense peak was at 21,269 Da.

protein with a molecular mass of 21,457 Da, as measured by mass spectrometry (Fig. 31). RhaR-NTD-His_6 protein contains four methionines, and if all four methionines were replaced by seleno-methionine, the molecular mass of the protein would be 21,457 Da, which precisely matched the mass of the purified protein, indicating that the seleno-methionine incorporation was successful.

The seleno-methionine labeled RhaR-NTD protein was concentrated to ~5 mg/ml, and used for crystallization. Crystals were obtained using the same conditions described above to obtain larger crystals, and were then soaked with L-rhamnose, transferred to cryoprotectant and sent to the Stanford Synchrotron Radiation Lightsource (SSRL) for X-ray diffraction.

Structure of RhaR-NTD-His_6 in complex with L-rhamonse. We have solved the crystal structure of the RhaR-NTD-His_6 in complex with L-rhamnose to 2.05 Å resolution (Fig. 32 and Table 11). The structure contained two protomers of RhaR-NTD in the asymmetric unit, with residues 6 to 165 in protomer A and residues 5 to 165 in protomer B of the dimer. The C-terminal His_6 and the flexible linker region between RhaR-NTD and His_6 tag were not visible, and protein residue numbers given here were numbering of RhaR-NTD only. The binding pockets in both protomers were fully occupied with L-rhamnose. The N- and C-terminal residues of the overexpression clone were not visible (residues 2-5 or 2-4 in protomers A and B, respectively, and 166-172 in both protomers), likely indicated that those regions were disordered. Each protomer of RhaR-NTD in the asymmetric unit was comprised of eight antiparallel β-strands (β1 to β8), four α-helices (α1 to α4), and a 3_{10} helix. The β-strands formed a cupin-superfamily β-barrel [138] containing the L-rhamnose-binding pocket. The pocket was enclosed by residues L7, F12, A23, R25, Q28, F31, H34, H36, F38, E40, H74, Y76, N86, and I88 (Fig. 33). The volume of this pocket was 230.8 $Å^3$, as calculated by CASTp [139], and contained one

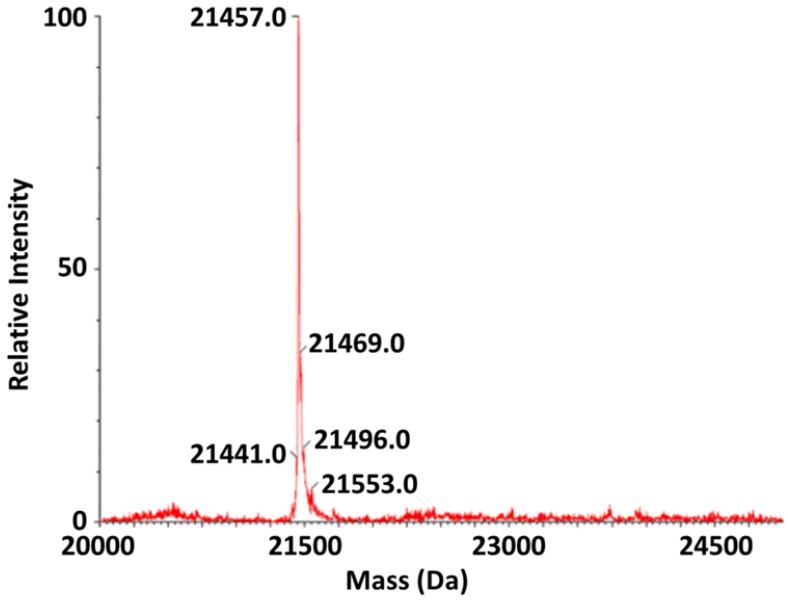


Fig. 31. The molecular mass profile of RhaR-NTD. Protein was overexpressed using a protocol modified form New England Biolabs (Ipswich, MA) to incorporate seleno-methionine. The most intense peak is at 21457 Da.

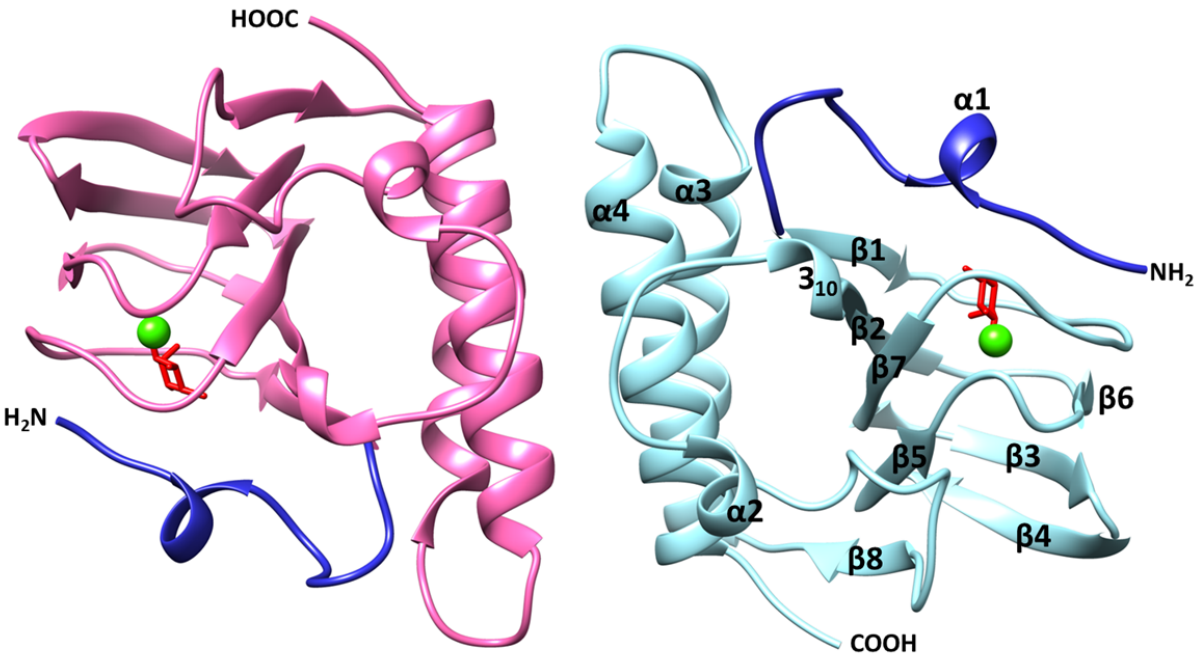


Fig. 32. The structure of RhaR-NTD dimer in complex with L-rhamnose. Protomer A is in magenta, protomer B is in cyan, L-rhamnose is in red stick, metal ion is in green sphere. The N- and C-termini, as well as secondary-structure elements are indicated. The arm region of each promotmer is colored blue. Protein structure figure was generated by UCSF Chimera package [140].

Table 11 Data collection and refinement statistics for RhaR NTD structure

Beamline	**SSRL 12-2**
Resolution/Å	**2.05(2.09-2.05)**
Completeness(%)	**99.8(98.9)**
I/σ	**31.7(7.78)**
R_{merge} †	**0.034(0.203)**
R_{work} ‡/R_{free} §	**0.181/0.213**
RMSD, Bond lengths/Å	**0.008**
RMSD, Bond angles/°	**1.15**
Overall B value (Å^2)	**38.06**
B value (Å^2) for Rhamnose	**31.83**

Values in parentheses are for the highest resolution shell.

†$R_{merge} = \sum |I - \langle I \rangle| / \sum I$, where I is measured intensity for reflections with indices hkl.

‡$R_{working} = \sum ||Fobs| - |Fcalc|| / \sum |Fobs|$.

§R_{free} has the same formula as $R_{working}$ except that calculation was made with the structure factors from the test set

molecule of L-rhamnose and one metal ion in each protomer. The metal ion was coordinated by residues H34, H36, E40, and H74 as well as two of the hydroxyl groups (-OH) of L-rhamnose.

A search for proteins with related structures (DALI [141]) identified the various AraC-NTD structures as the top hits, with PDB:1XJA as the best match among AraC-NTD structures (z=14, rmsd=2.5, 15% sequence identity). The structure of ExsA-NTD and ToxT were the second and third best AraC family protein hits (PDB: 4ZUA, z=10.4, rmsd=2.9, 10% sequence identity; and PDB:3GBG, z=7.7, rmsd=4.0, 7% sequence identity, respectively). The remaining structures related to RhaR-NTD were predominantly cupin β-barrel containing enzymes and seed storage proteins, including the bacilysin biosynthetic enzyme BacB structure (PDB: 3H7J), which was more closely related to the RhaR-NTD structure than ToxT.

Despite low sequence similarity, the overall fold of RhaR-NTD closely resembled those of the regulatory domains of AraC [33], ExsA [58], and ToxT [38], which all share a cupin-superfamily β-barrel. Overlays of the RhaR-NTD with the regulatory domains of AraC, ExsA, and ToxT are shown in Fig. 34. Structural alignments of the RhaR-NTD with AraC-NTD (PDB: 2ARC) and ExsA-NTD (PDB: 4ZUA) produces an r.m.s.d. of 2.6 Å and 2.9 Å (DALI [141]) respectively, with only 18% sequence identity. Superimposing RhaR-NTD with ToxT-NTD (PDB: 3GBG) produces an r.m.s.d of 3.4 Å (DALI [141]) with only 11% identical residues. The RhaR-NTD has one less β-strand than AraC, ExsA, and ToxT; the sequence that aligns with β2 in AraC, ExsA, and ToxT is an unstructured loop in the RhaR-NTD structure. The β-barrel structure of RhaR-NTD is aligned most closely with the AraC-NTD, with an r.m.s.d. of 1.84 Å (DALI [141]). The structural alignment of the ExsA and ToxT β-barrels, compared to the RhaR-NTD, produce an r.m.s.d. of 2.01 Å and 2.37 Å (DALI [141]), respectively. Overall, the RhaR-NTD β-barrel structure has a higher structural similarity to AraC, ExsA, and ToxT than the entire

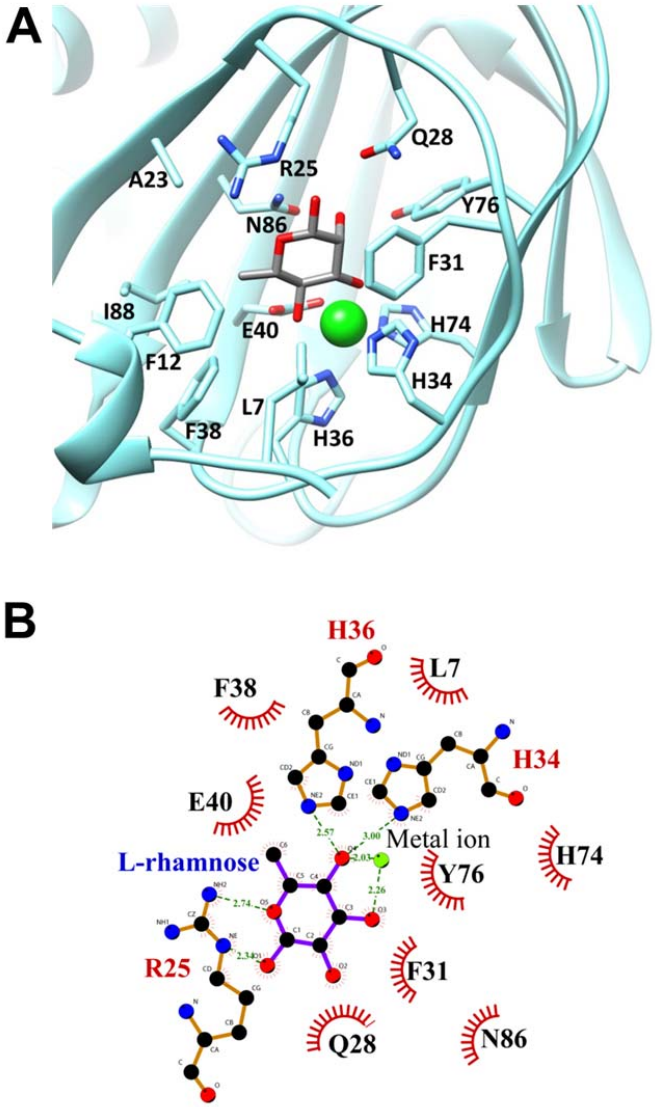


Fig. 33. RhaR L-rhamnose-binding site in detail. (A) Zoomed in view of the L-rhamnose binding pocket. L-rhamnose is in in gray stick, the metal ion is in green sphere. Residues enclose the rhamnose binding pocket are indicated. Protein structure figure was generated by UCSF Chimera package [140]. (B) A two-dimensional diagram showing interactions between RhaR-NTD and L-rhamnose as calculated by LIGPLOT[142]. Hydrogen bonds and distances are indicated in green and hydrophobic interactions are represented by an arc with spokes radiating towards the ligand atoms they contact.

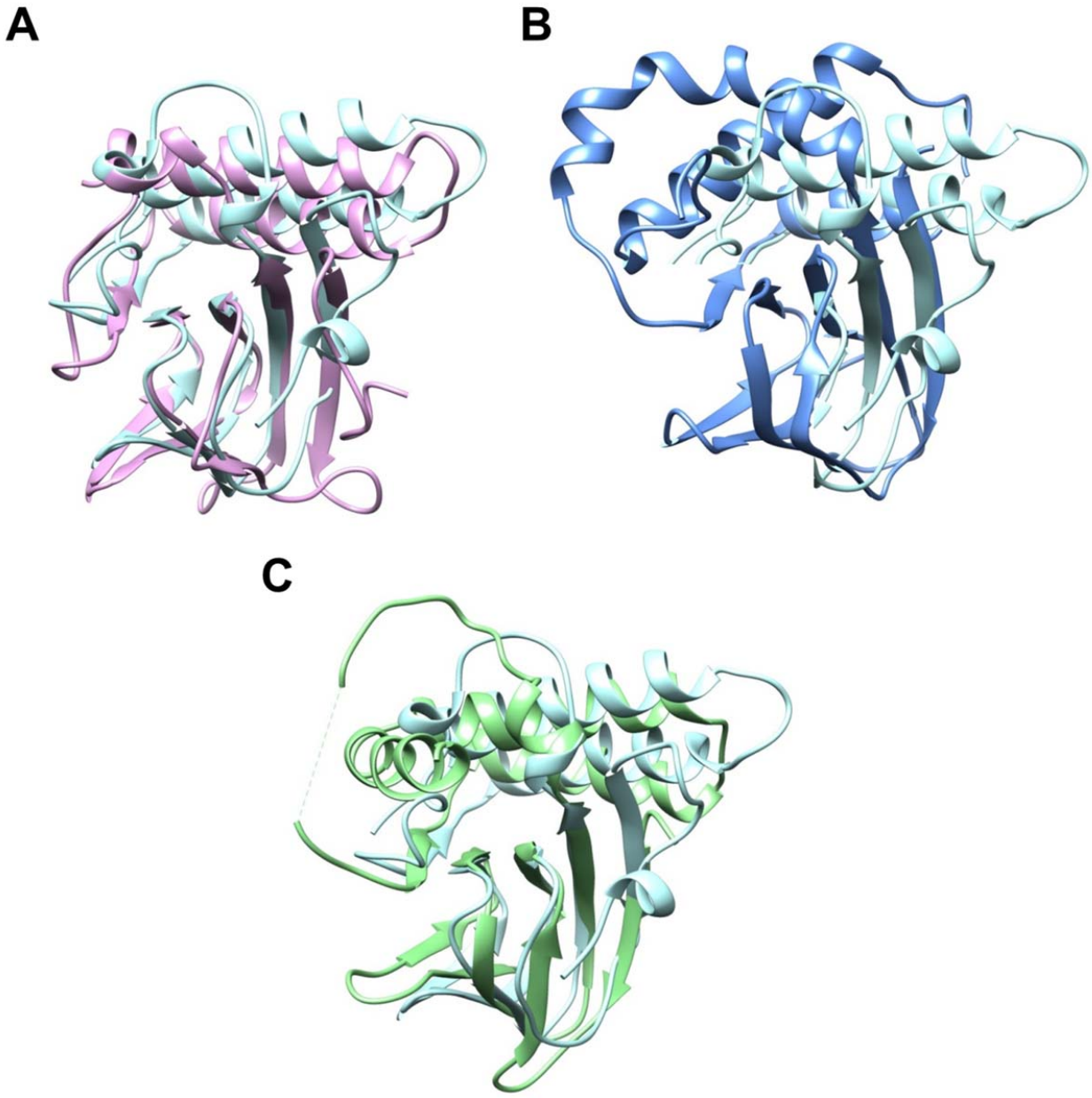


Fig. 34. RhaR-NTD overlays with AraC-NTD, ToxT-NTD, ExsA-NTD. RhaR-NTD is in cyan. (A) RhaR-NTD (light blue) superimposed with AraC-NTD. AraC-NTD: pink. **(B)** RhaR-NTD overlaid with ToxT-NTD. ToxT-NTD: blue. **(C)** RhaR-NTD structurally aligned with ExsA-NTD. ExsA-NTD: Green. Protein structure figure was generated by UCSF Chimera package [140].

structure does. Our results below suggest a biological rationale for the reduced structure in the region that aligns structurally with β2 in the other structures. Despite that the RhaR-NTD has 1 more helix than ExsA and 2 more than AraC; however, the helical contents of the RhaR-NTD aligns closely with those of the AraC-NTD and ExsA-NTD with an r.m.s.d. of 1.67 Å and 1.68 Å (DALI [141]), respectively. This structural similarity is mostly due to the similarity of α3 and α4 in RhaR-NTD to α2 and α3 in ExsA-NTD and in AraC-NTD. However, the overlay of the RhaR-NTD helices and ToxT-NTD helices gave a relatively large r.m.s.d. of 2.68 Å (DALI [141]), which is mostly because ToxT only contain one analogous helix α3 with RhaR-NTD.

The N-terminal 20 residues of AraC, prior to the first β-strand, are referred to as the arm. The AraC arm contains no secondary structure, however, conformational changes in this region are key to the mechanism that regulates AraC protein activity in response to arabinose [27, 29-31, 33, 143, 144]. The N-terminal arm of RhaR in our structure, as defined by residues before the first β-strand, consists of residues L5-A19. In contrast to the absence of secondary structure in the AraC arm, RhaR residues K9 to F12 formed a short α-helix (Fig. 35).

RhaR-NTD dimerization. In the presence of rhamnose, RhaR-NTD-His_6 crystallized with a dimer in the asymmetric unit. Similar to AraC and ExsA, the two protomers of RhaR-NTD in the crystals were in an antiparallel conformation, with an interface area of 890 $Å^2$. At each end of the dimerization interface, residues E136 and R161 (from helices α3 and α4, respectively) formed two hydrogen bonds, a salt bridge and multiple non-bonded interactions with each other (Fig. 36A). R161 from chain A also forms two hydrogen bonds with P142 from chain B, and R161 from both chains forms multiple additional non-bonded interactions. Near the center of the dimerization interface, the side chain of Q154 from protomer B forms two hydrogen bonds, one

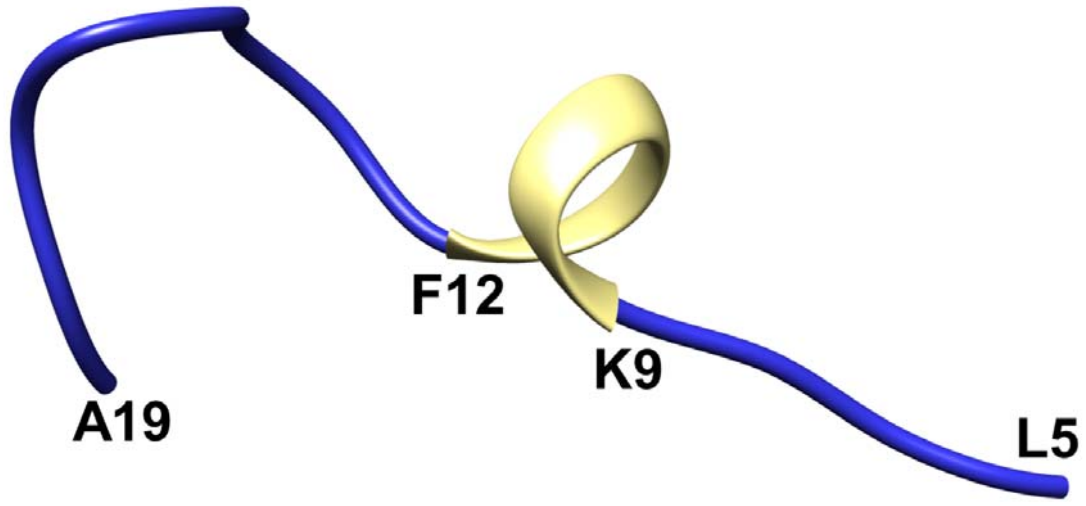


Fig. 35. Zoomed in view of the RhaR-NTD arm region (residues L5 to A19). The α helix region (residues K9 to F12) was highlighted in yellow. Protein structure figure was generated by UCSF Chimera package [140].

each to Q132 and Q154 from protomer A. Residues Q132, E136, Q154 and R161 are each ≥95% identical among RhaR protein sequences, consistent with the importance of these residues in RhaR dimerization. In AraC, the dimerization interface is anchored at the ends by a triad of leucine residues, providing stability [33]. Although leucine triads aren't found in RhaR-NTD, nine additional residues in chain A and eleven additional residues in chain B are predicted to participate in non-bonded interactions, contributing to dimerization strength. A detailed schematic illustrating the interactions between the protomers is shown in Fig. 37.

In addition to the contacts between residues in α3 and α4, residues K95, L96, N97 and L98 in the loop between α2 and the 3_{10} helix from each protomer are positioned to make contacts that likely provide additional dimer stability (Fig. 36B); in AraC, residues R101 and Y103 in the analogous region are positioned to make similar interactions [33]. In RhaR-NTD, N97 in each chain can form two hydrogen bonds with N97 and one hydrogen bond with K95 in the opposite chain. Further, the residues in these positions are identical (L96 and N97) or conserved as positively charged (K95) or hydrophobic (L98) in ≥89% of RhaR protein sequences.

L-rhamnose binding pocket. Each protomer of RhaR-NTD contained a molecule of L-rhamnose and a metal ion within the open end of a small pocket formed by the cupin fold (Fig. 30). This rhamnose binding pocket is much smaller in volume (230.8 $Å^3$ [139]) than the sugar binding pocket of AraC (401.8 $Å^3$ [139]) or the fatty acid binding pocket of ToxT (789.9 $Å^3$ [139]), and there are no water molecules found within the RhaR-NTD pocket. In AraC, an extensive network of bound water molecules in the arabinose-binding pocket mediates hydrogen bonding between the arabinose and AraC [33]. In RhaR, residues R25, Q28, and N86 are positioned to directly interact with rhamnose through hydrogen or electrostatic bonding; and F31 is positioned so that its benzyl ring could stack with rhamnose. We constructed alanine

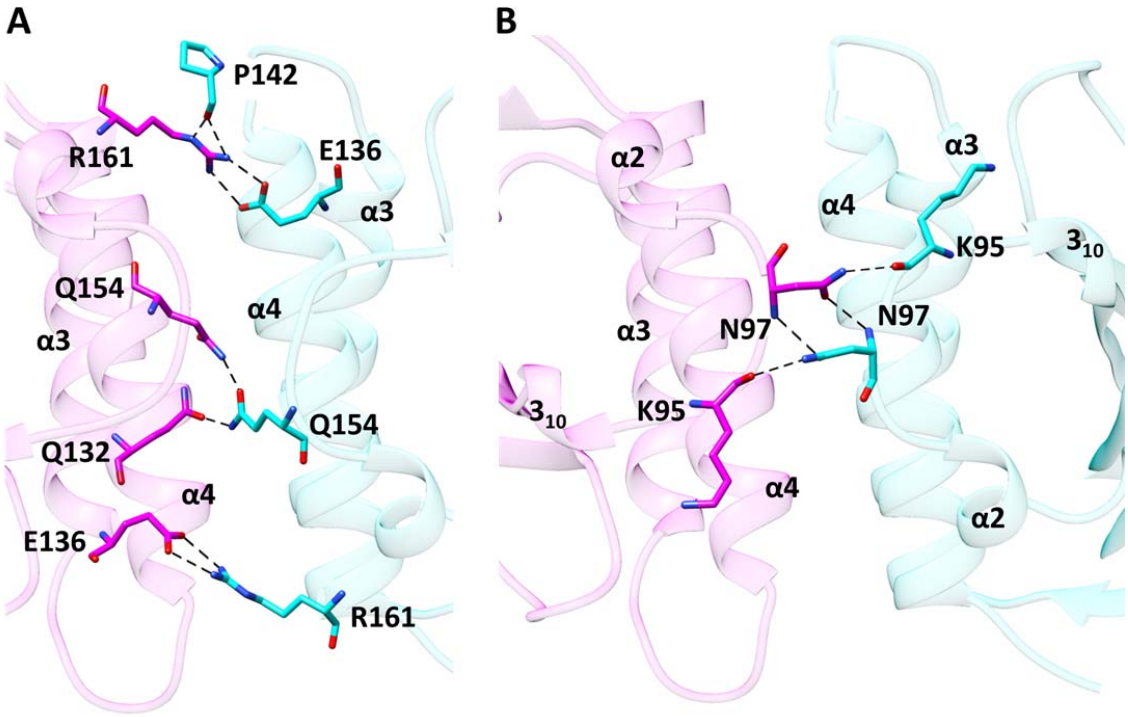


Fig. 36. **(A)** Hydrogen-bond (H-bond) interactions at the RhaR-NTD dimerization interface between protomer A and B. **(B)** H-bond interactions at the loop connecting α2 and the 3_{10} helix between protomer A and B. Protomer A is in magenta and protomer B is in cyan. Helices α2,α3,α4, and 3_{10} are labeled. Residues involved in making H-bond interactions at the interface are indicated in the same color as each of the protomer backbone. H-bonds are highlighted in black dash lines. Protein structure figure was generated by UCSF Chimera package [140].

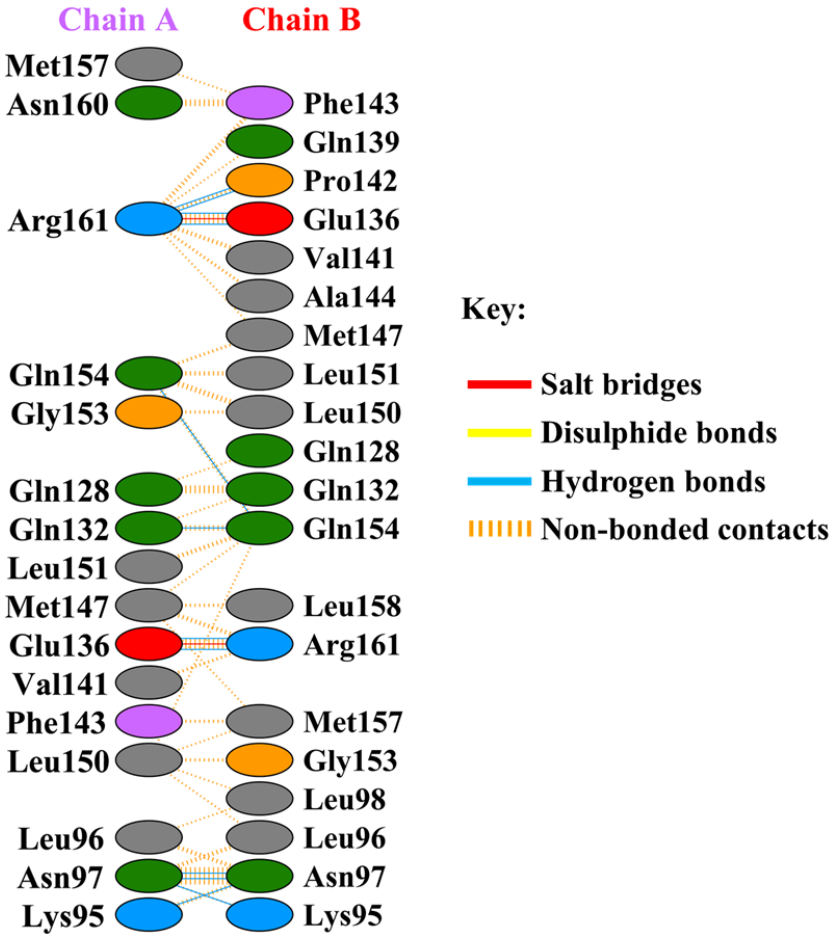


Fig. 37. Schematic diagram of interactions between RhaR protomers. Residues are colored by their type. Positively charged residues (His, Lys, Arg) are blue; negatively charged residues (Asp, Glu) are red; neutral residues (Ser, Thr, Asn, Gln) are green; aliphatic residues (Ala, Val, Leu, Ile, Met) are grey; aromatic residues (Phe, Tyr, Trp) are purple; proline and glycine residues are orange. Interactions are indicated by lines between residues. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, the width of the striped line is proportional to the number of the atomic interactions. Diagram generated by PDBsum [145-148].

substitutions at positions R25, Q28, and N86, and found that transcription activation by these variants were down by 25 to 30% in the presence of rhamnose [(+)rha] compared to wild-type RhaR (Fig. 38A), supporting the hypothesis that these residues contribute to the strength of rhamnose binding. We would not have predicted that in the absence of rhamnose [(-)rha], Q28A and N86A both would showed slightly increased activity, and R25A would have approximately 2-fold increased activity relative to wild-type RhaR. An F31A variant of RhaR-NTD was also constructed. Transcription activation by this variant was approximately the same as wild type (+)rha, however, the variant exhibited a 2.5-fold increase in activity (-)rha compared to wild type (Fig. 38A). The wild-type activity of the F31A variant (+)rha suggests that rhamnose stacking with the aromatic ring of F31 does not contribute as much to rhamnose binding in RhaR as the stacking between arabinose and the aromatic ring of W95 in AraC [33, 149]. The finding that both R25A and F31A had 2- to 2.5-fold increased activity (-)rha compared with wild type suggests that these residues play roles in the (-)rha conformation, possibly by helping to maintain a (-)rha conformation that has reduced activity relative to the (+)rha conformation. Lowden *et al* have proposed that interactions (bridged by the fatty acid effector) between ToxT residues K31 within the effector binding pocket and K230 in the DBD function to hold ToxT in its non-activating (reduced activity) conformation [38]. This suggests the hypothesis that one or both of RhaR R25 and F31 may contact RhaR DBD in the absence of rhamnose to hold RhaR in its lower activity (-)rha conformation. Results below provide evidence that a region including these residues undergoes a rhamnose-dependent conformational change, consistent with this hypothesis.

Interestingly, the bound rhamnose also interacts with a metal ion that is positioned deep within the rhamnose-binding pocket, suggesting the metal ion may contribute to rhamnose

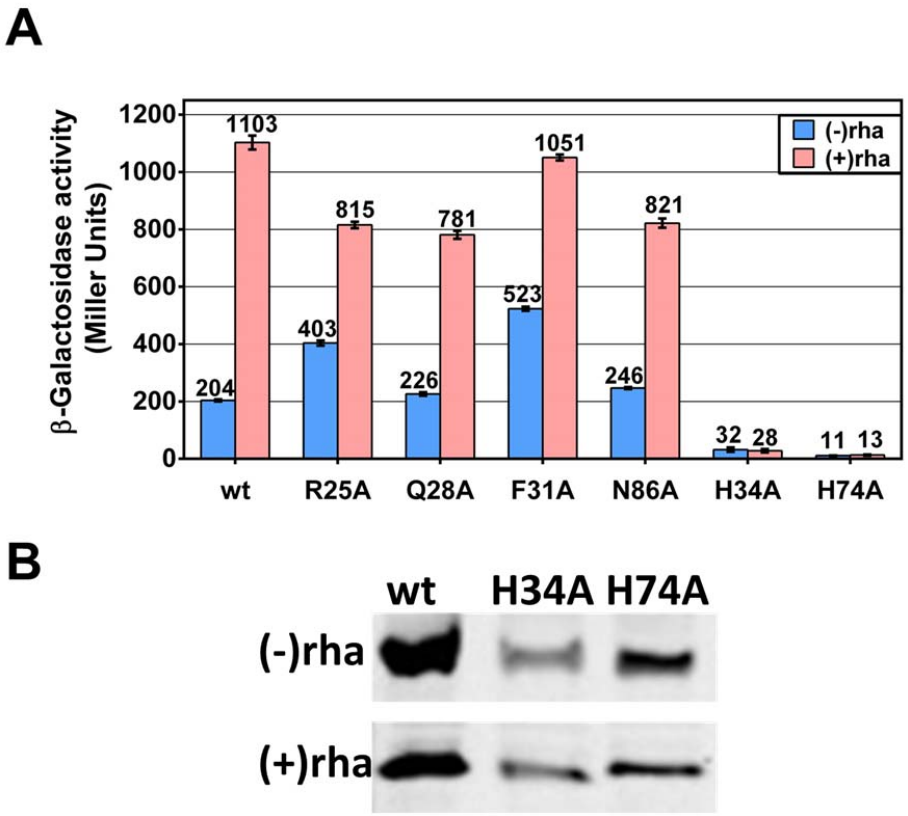


Fig. 38. **(A)** Transcription activation by RhaR variants. Plasmids expressing wild-type RhaR or variants were transformed into strain SME2525 [Φ(*rhaS-lacZ*)Δ128 Δ*rhaSR*::*kan recA*::*cat*], the cells were grown with or without L-rhamnose, as indicated, and β-galactosidase activity was measured in Miller units. **(B)** Western blots showing the protein levels of the RhaR variants. The culture samples in all of the lanes were loaded at equivalent cell densities (based on OD600).

binding. The metal ion is located in the center of the cupin barrel of each domain. The metal ion is coordinated by four residues: H34 and H36, which are on a loop connecting β1 and β2; E40 on β2; and H74 on β6. These four residues are all highly conserved among proteins in the cupin superfamily of proteins, function to coordinate metal ions [150]. These residues are 100% conserved among RhaR protein sequences, suggesting that metal binding are conserved among RhaR proteins. The total coordination number of the metal ion is six: the three His residues each coordinate via their NE2 atoms, the Glu residue via OE2, and rhamnose via O3 and O4. The metal ion-ligand distances are between 1.99 and 2.27 Å (Fig.39).

To further examine the role of the His residues within the sugar-binding pocket, we constructed individual alanine substitutions at H34 and H74. In both of these variants, transcription activation was substantially decreased both in the presence [(+)rha] and the absence of rhamnose [(-)rha] as compared to the wild type, and both of these variants showed no increasein activation (+)rha relative to (-)rha (Fig. 38A). Western blot data (Fig. 38B) showed that the protein levels of H34A and H74A were decreased by 75% and 60%, respectively, relative to wild-type RhaR both (-) and (+)rha, suggesting a decrease in protein stability. It is likely that the instability of the H34A and H74A variants was at least largely the result of the loss of metal binding; however, conformational changes in the binding pocket due to the alanine substitutions may also contribute.

Structure of rhamnose-free RhaR-NTD-His$_6$. In order to obtain rhamnose-free RhaR-NTD crystals, RhaR-NTD-His$_6$ protein was over-expressed in a M9 salt minimal media [151], purified, and crystallized without addition of rhamnose. Two structures of the rhamnose-free RhaR-NTD were determined. One structure diffracted to 1.7 Å resolution, and contained an ethylene glycol

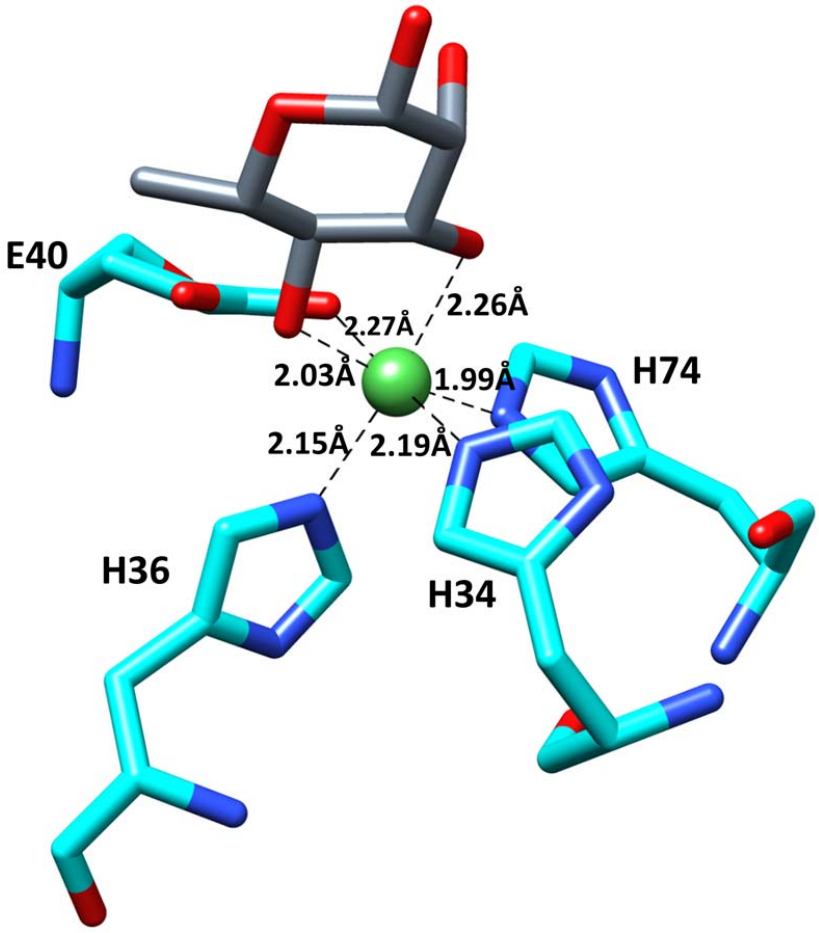


Fig. 39. Metal ion coordination in detail. The protein ligands are His34, H36, H74, and E40. A L-rhamnose molecule completes the coordination sphere. The metal ion is drawn in green sphere, and the metal ion-ligand distances are indicated. Protein structure figure and analysis were performed with UCSF Chimera package [140].

molecule within the sugar binding pocket of each protomer (RhaR-NTD-EG) (Fig. 40A, Table 12). The other structure was diffracted to 1.9 Å resolution, and a glycerol molecule was found within the sugar binding pocket of each protomer (RhaR-NTD-GLY) (Fig. 40B, Table 12). Overall, the two rhamnose-free RhaR-NTD structures are nearly identical with an RMSD of 0.33 Å between Cα atoms following superposition of the dimers with GESAMT (General Efficient Structural Alignment of Macromolecular Targets) [152].

A metal ion was also present within the sugar-binding pocket in each protomer. Given that the crystallization solution contains 50 mM $Ca(OAc)_2$, RhaR-NTD could acquire the Ca^{2+} from the crystallization solution since it was most abundant. However, refinement of the RhaR-NTD structure with Ca^{2+} ion resulted in a large amount of residual positive electron density around the ion, which suggested that a larger metal was bound within the sugar-binding pocket. Furthermore, the metal ion-ligand distant was approximately 2.0 Å to 2.1 Å, which was much shorter than the expected distance of ~2.4 Å for a Ca^{2+} ion [153]. The phased anomalous difference map calculated using data collected at a wavelength of 1.0000 Å yielded peaks with heights 38σ and 30σ at the rhamnose-bind pocket in protomers A and B, respectively. This indicated that a larger ion (for instance, Zn^{2+}) that would give a strong anomalous signal at λ=1.0000 Å occupied these sites. However, an X-ray fluorescence scan of the RhaR-NTD crystals revealed that Ni^{2+} ions were present (Fig. 41A). An additional X-ray fluorescence scan was carried out to determine whether Zn^{2+} could also be present. However, no zinc signal was detected (Figure 41B). Additional X-ray diffraction data were collect for RhaR-NTD at wavelengths of 1.4848 Å (8,350 eV) and 1.4938 Å (8,300 eV) which corresponded to the emission peak and low energy remote respectively. The phased anomalous difference maps

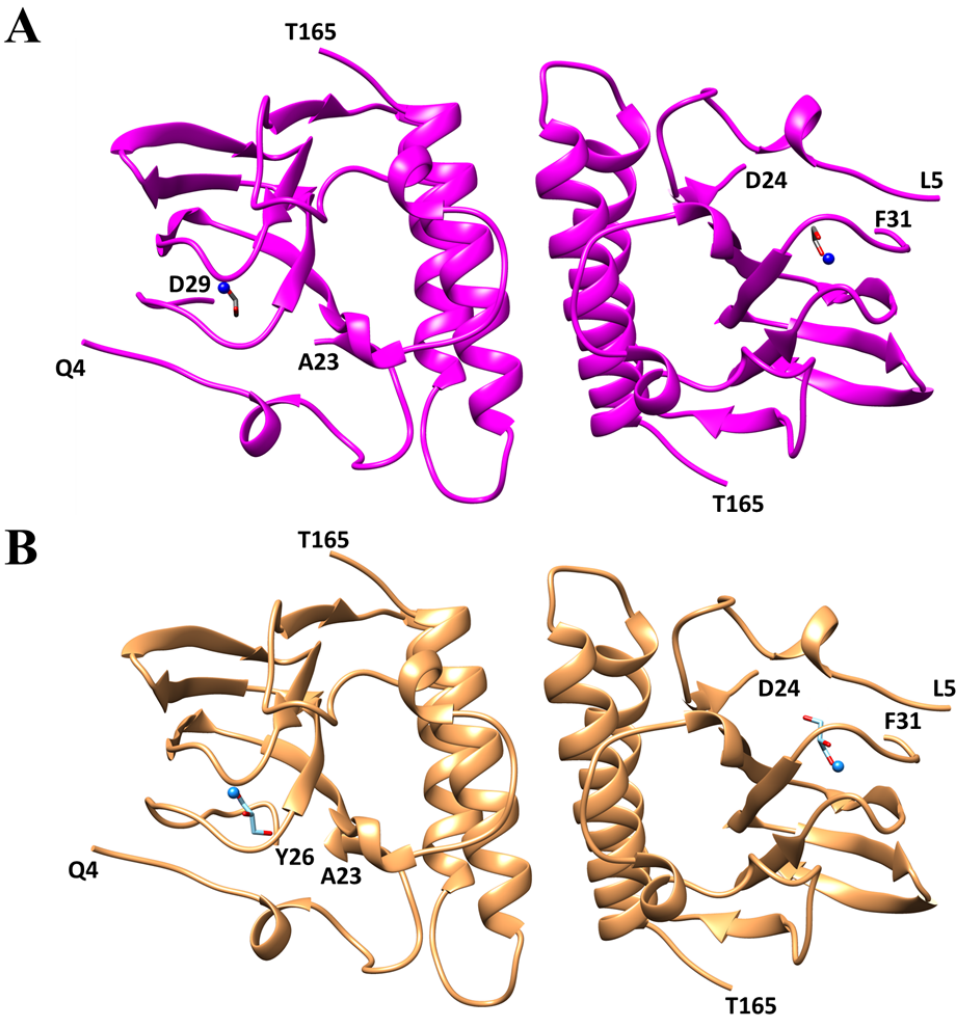


Fig. 40. Structures of rhamnose-free RhaR-NTD-EG and RhaR-GLY. (A) 1.73 Å RhaR-NTD-EG structure. Ethylene glycol molecules are in sticks, Ni^{2+} ions are in dark blue sphere. **(B)** 1.90 Å RhaR-NTD-GLY structure. Glycerol molecules are in sticks, Ni^{2+} ions are in light blue sphere.

Table 12. Crystallographic data for apo RhaR-NTD.

	RhaR-NTD-EG	RhaR-NTD-GLY
Data Collection		
Unit-cell parameters (Å, °)	a=b=101.82, c=96.14	a=b=101.64, c=96.84
Space group	$P4_32_12$	$P4_32_12$
Resolution (Å)[1]	48.07-1.73 (1.76-1.73)	48.42-1.90 (1.94-1.90)
Wavelength (Å)	1.0000	1.0000
Temperature (K)	100	100
Observed reflections	703,229	529,996
Unique reflections	53,302	40,601
$<I/\sigma(I)>$[1]	19.3 (1.7)	19.4 (1.8)
Completeness (%)[1]	100 (100)	100 (100)
Multiplicity[1]	13.2 (13.6)	13.1 (13.1)
R_{merge} (%)[1, 2]	7.6 (177.9)	8.1 (164.4)
R_{meas} (%)[1, 4]	7.9 (184.9)	8.4 (170.9)
R_{pim} (%)[1, 4]	2.2 (50.0)	2.3 (47.0)
$CC_{1/2}$ [1, 5]	0.999 (0.634)	0.999 (0.669)
Refinement		
Resolution (Å) [1]	36.00-1.73	35.94-1.90
Reflections (working/test)[1]	50,538/2,687	38,573/1,956
R_{factor} / R_{free} (%)[1,3]	17.5/20.0	17.9/20.1
No. of atoms (Protein/ Ni^{2+}/Ethylene glycol or glycerol/Water)	2,500/2/12/188	2,527/2/18/144
Model Quality		
R.m.s deviations		
Bond lengths (Å)	0.010	0.011
Bond angles (°)	0.974	1.011
Average B-factor (Å^2)		
All Atoms	40.1	49.6
Protein	39.7	49.4
Ni^{2+}	32.2	38.5
Ethylene glycol or glycerol	42.3	60.8
Water	45.9	50.8
Coordinate error(maximum likelihood) (Å)	0.21	0.20
Ramachandran Plot		
Most favored (%)	98.4	97.4
Additionally allowed (%)	1.6	2.3

1) Values in parenthesis are for the highest resolution shell.

2) $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the ith reflection and $<I(hkl)>$ is the average intensity of all reflections with indices hkl.

3) $R_{factor} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$; Rfree is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

4) R_{meas} = redundancy-independent (multiplicity-weighted) R_{merge} [92, 122]. R_{pim} = precision-indicating (multiplicity-weighted) R_{merge} [123, 124].

5) $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data [125, 126].

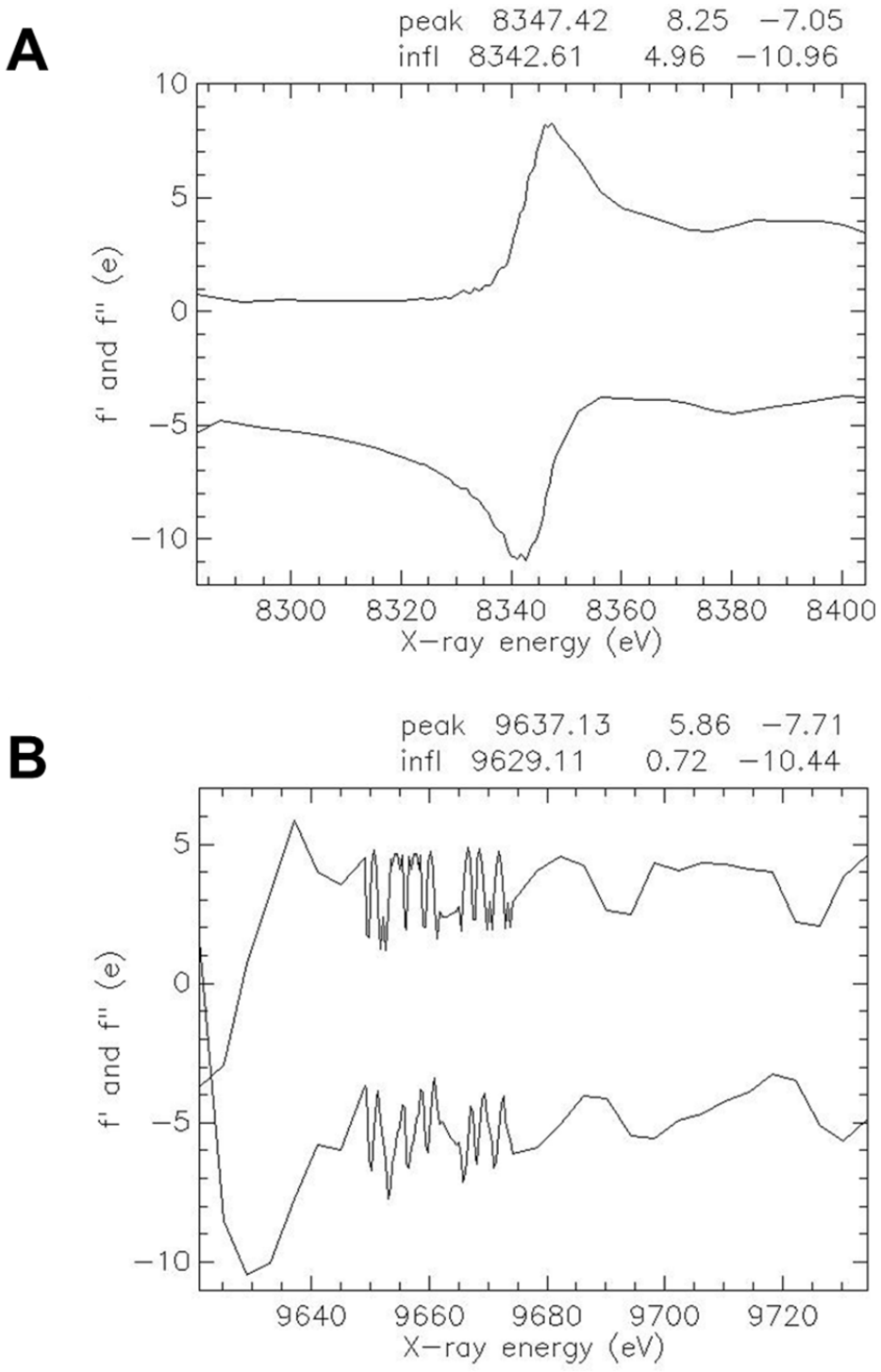


Fig.41. X-ray fluorescence spectra from a rhamnose-free RhaR-NTD crystal. (**A**) Nickel emission and (**B**) Zinc emission.

calculated using the peak and low energy data sets yielded peaks in protomers A/B of 44σ/35σ (peak wavelength) and 5.6σ/5.7σ (low energy wavelength) thus confirming the Ni^{2+} was the metal ion bound to RhaR-NTD. Since the RhaR-NTD protein was purified through Ni^{2+} affinity chromatography, it is likely that the Ni^{2+} was acquired during the purification. Throughout all the purification steps, the Ni^{2+} was the only divalent metal the protein was exposed to. It was interesting that RhaR-NTD retained the bound Ni^{2+} through the size exclusion chromatography purification step, and the high concentration of Ca^{2+} in the crystallization solution did not replace the Ni^{2+}. These all suggested that the metal binding to the protein was tight. Although X-ray fluorescence scan was not performed on the rhamnose-bound RhaR-NTD crystals, it is reasonable to believe that the metal ion present in the rhamnose-bound RhaR-NTD structure were also Ni^{2+} because the protein used to produce these crystals was also purified through the Ni^{2+} affinity chromatography. The refinement of the rhamnose-bound RhaR-NTD structure with Ni^{2+} did not yield any large amount of residual positive electron density around the ion, which suggested that Ni^{2+} fitted well within the sugar-binding pocket. In addition, the metal-ion distance was approximately 1.99 Å to 2.27 Å, which was also the expected distance of a Ni^{2+} ion (1.99 Å to 2.30 Å) [154].

The Ni^{2+} ion is coordinated by residues H34, H36, H74 and E40, same as the rhamnose-bound RhaR-NTD structure mentioned above. Following refinement with the Ni^{2+} ions placed in each protomer, large difference electron density peaks were present near the "open" coordination site of the metal ions. For the RhaR-NTD-GLY structure, the electron density was consistent with glycerol molecules coordinating the Ni^{2+} ions (Fig. 42A). The glycerol was likely acquired from the protein storage buffer (5%, v/v). However for the RhaR-NTD-EG structure, the glycerol at the metal coordination site was displaced by ethylene glycol during cryoprotection

(Fig. 42B). It should be noted that data collected using crystals cryoprotected with PEG 200 and PEG 400 displayed electron density consistent with a glycerol molecule, similar to RhaR-NTD-GLY, at the metal coordination site.

Overall, the rhamnose-free RhaR NTD-EG structure is similar to the rhamnose-bound structure with RMSD of 0.7 Å for the backbone C_α trace (Fig. 43A). In particular, two of the helices α4 and α5 that are involved in dimerization are well aligned in the two structures, suggesting that the dimerization is likely unaffected by rhamnose binding. However, there are also several differences between the two structures, with the largest difference being the positions of the backbones in the region from residue V22 to residue T35 (Fig 43B). Within this region, residues D24 to Q28 in protomer A, and residues R25 to V30 in protomer B of the rhamnose-free structure are completely disordered. It is clear that the position of the backbone in the region from D29 to T35 (in protomer A, F31 to T35 in protomer B) in the rhamnose-free structure has changed as compared to the rhamnose-bound structure (Fig. 43B). Residues R25, Q28, F31, and H34 are positioned to contact rhamnose, thus the lack of these interactions when rhamnose isn't bound has caused region D24 to Q28 (in protomer A, R25 to V30 in protomer B) became more flexible in the isolated RhaR-NTD structure. The V22 to T35 region in RhaR aligns with a region in ToxT that contacts the ToxT DBD, suggesting this region's potential involvement in transmitting the rhamnose binding signal to the DBD. It is likely that this region contacts the DBD in the absence of rhamnose, and changes conformation so that several residues (R25, Q28, F31 and H34) are in position to interact with the rhamnose when rhamnose becomes available, transmitting the allosteric signal from the NTD to the DBD. In addition, a region of β6 and the loop connecting β6 and β7 (residues S75 to V83) also shows backbone

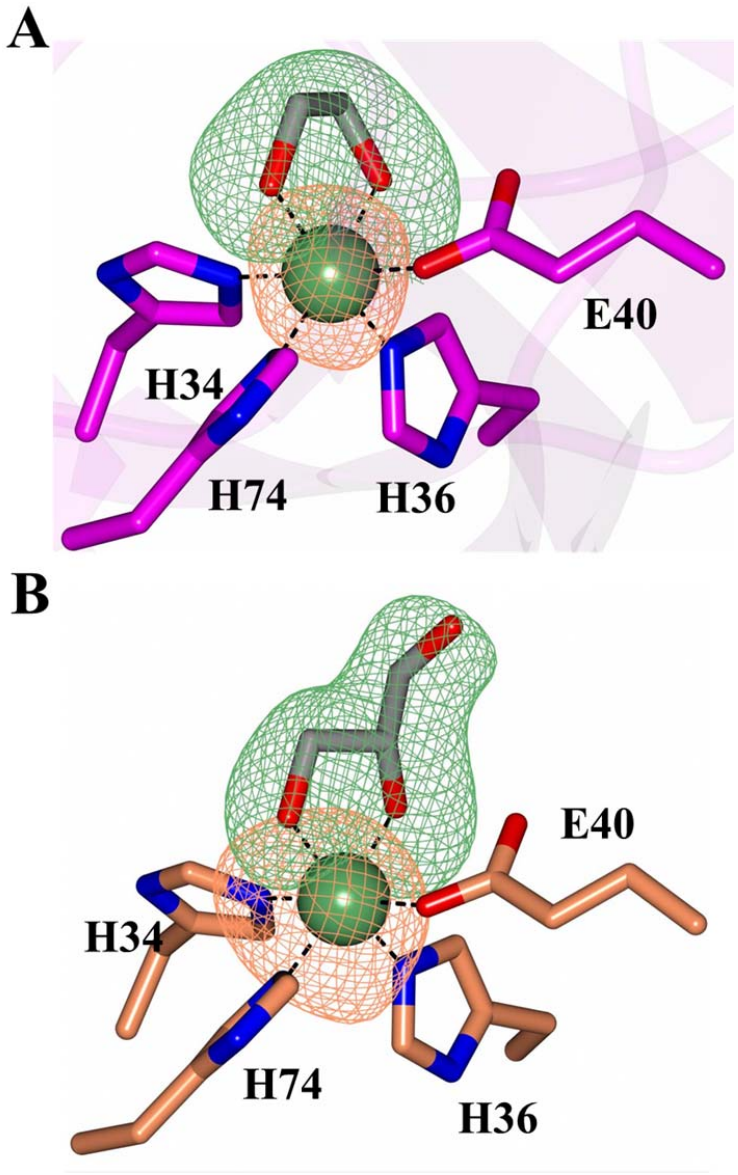


Fig. 42. Fo-Fc omit map (green mesh) and phased anomalous difference map (orange mesh) contoured at 3σ. (A) RhaR-NTD-EG and **(B)** RhaR-NTD-GLY. The anomalous difference maps were calculated using data collected at 1.0000 Å and 1.4848 Å for RhaR-NTD-EG and RhaR-NTD-GLY respectively. The Ni^{2+} ions are drawn as green spheres.

movement (Fig. 43C). Residue S75 to N80 forms a loop connecting β6 and β7 when rhamnose is bound; however, in the rhamnose-free structure, these residues became part of the β6. The conformational changes of the region from residue S75 to V83 suggests that this region may also involve in transmitting the rhamnose binding signal from the NTD to the DBD.

The N-terminal arm regions (residues L4-A19) of the rhamnose-free RhaR-NTD are well aligned with the rhamnose-bound structure, suggesting a lack of conformational change between the (+)rha and (-)rha states. In AraC, the analogous N-terminal arm region is key to allosteric arabinose signalling: the arm contacts the DBD in the absence of arabinose, constraining AraC in its non-activating state; in the presence of arabinose, the arm adopts an alternative conformation and releases the DBD, thus allowing AraC to activate transcription of *araBAD* (the 'light-switch' mechanism) [35]. Our previous N-terminal deletions of the arm region of RhaR [52] did not have rhamnose-dependent effects on RhaR activity, arguing that RhaR likely does not use amechanism analogous to the AraC 'light-switch' mechanism. Instead, deletions of the RhaR arm had similar effects on the (+)rha and (-)rha states, consistent with the lack of conformational differences in the structures [52]. The arm region in AraC folds over the sugar-binding pocket in the presence of arabinose but lays on one edge of the sugar-binding pocket rim in the absence of arabinose. The RhaR arm region, on the other hand, stays on one edge of the rhamnose-binding pocket rim regardless weather rhamnose is bound or not. This rim of the pocket is the opposite edge of the pocket rim that shows the conformational change in RhaR.

Overall, the conformational changes between the (+) and (-)rha conformations are relatively small; however, prior studies suggest that this is reasonable. It has previously been shown that *in vivo* and *in vitro* RhaR activity increases by approximately 10-fold (+)rha relative to (-)rha [51, 53]. It seems plausible that this relatively small change in RhaR activity (RhaS

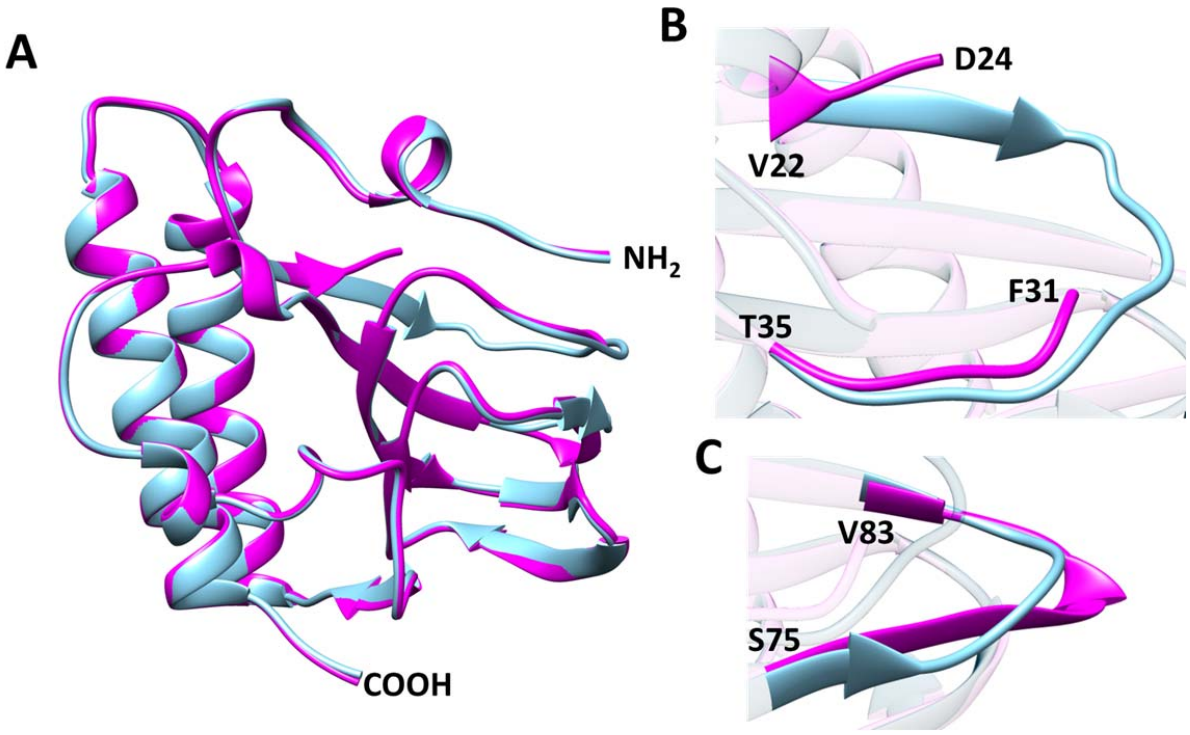


Fig. 43. **(A)** The superimposition of the rhamnose-free RhaR-NTD-EG protomer to the rhamnose-bound structure protomer. Cyan, rhamnose-bound RhaR-NTD structure; magenta, rhamnose-free RhaR-NTD-EG structure. The rhamnose, ethylene glycol, and Ni^{2+} are not shown. **(B)** Zoomed in view of the V22 to T35 region and **(C)** the region of S75 to V83 showing structure differences of these regions between rhamnose-bound and free RhaR-NTD structures.

activity increases by 1000x upon addition of rhamnose, for example) might be achieved through relatively modest conformational changes upon rhamnose binding. The finding that four residues in the V22-T35 loop, R25, Q28, F31, and H34, are positioned to contact rhamnose provides a potential mechanism for the conformational change in this region. Further, our finding that alanine substitutions at R25 and F31 in this loop had increased activity in the absence of rhamnose suggests that these residues promote the non-activating, lower activity state of RhaR in the absence of rhamnose. By analogy with ToxT, in which K31 interacts with K230 to stabilize the non-activating state of ToxT, R25 and F31 and potentially other residues in the V22-H34 loop may contact RhaR-DBD to stabilize the non-activating (-)rha state of RhaR.

Partially rhamnose-occupied RhaR-NTD structure. Prior to obtaining the structure of RhaR-NTD in the absence of rhamnose, we had made a prior effort to obtain a rhamnose-free RhaR-NTD structure. We carried out all of the steps to obtain purified protein (cell growth, protein purification, crystallization) in the absence of rhamnose. We solved the structure of RhaR-NTD from protein crystals grown from this protein at 2.02 Å resolution, however, the structure had electron density within the rhamnose-binding pocket. The finding that the electron density in the pocket had a higher average *B* factor than the surrounding protein atoms (Table 13) suggested the possibility that the pocket was only partially occupied with the ligand. Thus, we hypothesize that the density in the pocket was due to rhamnose binding to some, but not all, of the protein in the crystal. This partial occupancy is suggested by comparison of the $2F_o - F_c$ electron-density maps for the two structures (Fig. 44). The source of the rhamnose bound to RhaR-NTD in this structure is unknown, however the yeast extract that was a component of the growth medium seems a likely candidate. The finding that RhaR-NTD apparently retained the bound rhamnose

Table 13 Data collection and refinement statistics for RhaR NTD structure with partial L-rhamnose occupancy

Beamline	**APS**
Resolution/Å	**2.02(2.07-2.02)**
Completeness(%)	**99.6(95.8)**
I/σ	**20.1(3.0)**
$\boldsymbol{R}_{\mathbf{merge}}$ †	**0.049(0.502)**
$\mathbf{R}_{\mathbf{work}}$ ‡/$\mathbf{R}_{\mathbf{free}}$ §	**0.185/0.207**
RMSD, Bond lengths/Å	**0.008**
RMSD, Bond angles/°	**1.15**
Overall B value (Å^2)	**42.90**
B value (Å^2) for Rhamnose	**64.04**

Values in parentheses are for the highest resolution shell.

†R_{merge} =∑ |I − <I>|/∑ I, where I is measured intensity for reflections with indices hkl.

‡$R_{working}$ = ∑ ||Fobs| − |Fcalc||/ ∑| Fobs|.

§R_{free} has the same formula as $R_{working}$ except that calculation was made with the structure factors from the test set

through multiple chromatographic purification steps suggests that the binding was tight, with a low dissociation constant or at least a slow off rate for the rhamnose.

Overall, the partially rhamnose-occupied structure is similar to the fully rhamnose-occupied structure and the rhamnose-free structure, with an r.m.s.d of 0.7 Å, and 0.5 Å respectively for the backbone C_α trace (45A). The N-terminal arm region as well as the two α helices α4 and α5 involved in dimerization are well aligned both in backbone and side chains in the three structures. Similar to the rhamnose-free RhaR-NTD structure, residues R25 to D29 in protomer B of the partially rhamnose-occupied structure were completely disordered, but in protomer A, the backbone electron density could be traced although the side chain electron density was poor. Comparison of the three RhaR-NTD structures shows that two regions (V22 to H34 and S75 to V83) have the largest difference. The partial rhamnose-occupied structure region V22 to T35 (specifically V22 to A23, and V30 to T35) aligned well with the same region in the rhamnose-free structure. In the region of S75 to V83 in the partial rhamnose-occupied structure, S75 to A77 aligned with the same region in the rhamnose-free structure, and S78 to V83 aligned with the same region in the full rhamnose-bound structure (Fig. 45B & C). These observations support the idea that partially rhamnose-occupied structure represents an average of the (-) and (+)rha conformations of RhaR-NTD.

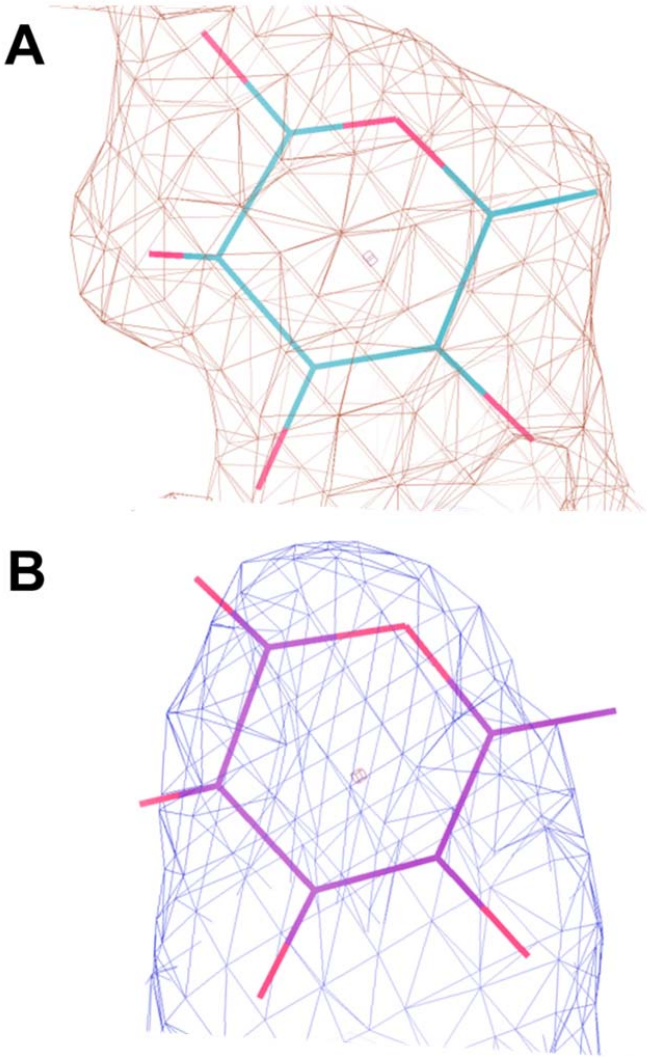


Fig. 44. 2*Fo* – *Fc* electron-density maps of rhamnose with full and partial occupancy (contoured at 1σ). **(A)** Full rhamnose occupancy map. **(B)** Partial rhamnose occupancy map.

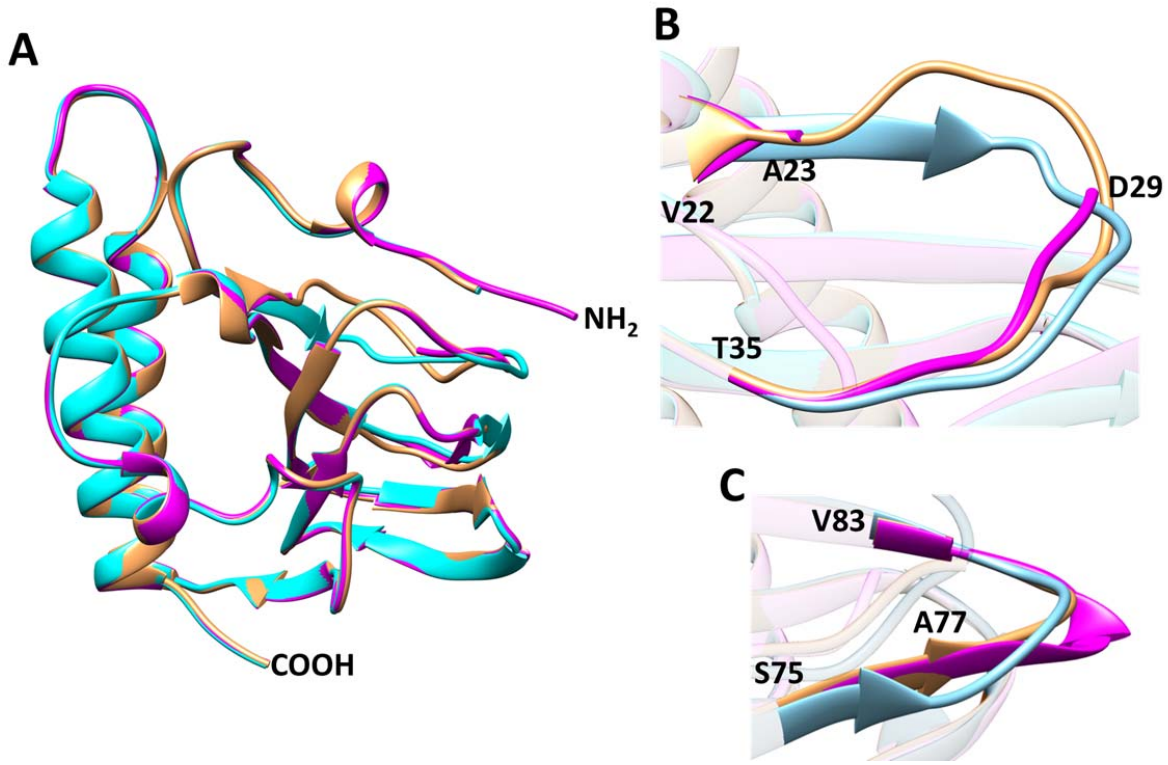


Fig. 45. (A) The structure overlay of the full rhamno **Structures of RhaR-NTD** se-bound RhaR-NTD, partial rhamnose-occupied RhaR-NTD, and rhamnose-free RhaR-NTD. Cyan, full rhamnose-bound RhaR-NTD structure; magenta, rhamnose-free RhaR-NTD-EG structure; brown, partial rhamnose-occupied RhaR-NTD. The rhamnose, ethylene glycol, and Ni^{2+} ions are not shown. **(B)** Zoomed in view of the V22 to T35 region and **(C)** the region of S75 to V83 showing backbone differences of these regions between the three RhaR-NTD structures.

Chapter 8

DNA binding and Allosteric signaling by RhaR

In the presence of L-rhamnose, the AraC family transcriptional activator RhaR increases transcription activation of the *rhaSR* operon by 10-fold relative to the absence of rhamnose [13] [10-12]. The mechanism of the increase in RhaR activity when it has L-rhamnose bound isn't fully understood. Prior studies have provided evidence that L- rhamnose binding to RhaR results in a 25-fold increase in DNA binding affinity, and (even if DNA binding is saturated) is required post-DNA binding for transcription activation (presumably through increasing contacts with the RNA polymerase sigma subunit) [48, 53, 54]. Although RhaR DNA-binding affinity isn't the only property that is affected by L-rhamnose binding, measurements of L-rhamnose-dependent effects on DNA binding by RhaR are expected to provide information about the mechanisms of allosteric L-rhamnose signalling in RhaR. Here, by using Electrophoretic mobility shift (EMSA) assays, I have measured the dissociation constants, a measurement of DNA-binding affinity, for the interactions between RhaR and the isolated RhaR-DBD with RhaR DNA binding sites.

Equilibrium binding measurements. Tobin and Schleif previously performed equilibrium DNA binding assays of RhaR binding to the full RhaR binding site upstream of the *rhaSR* promoter in the absence and presence of rhamnose and to one of the RhaR half-sites in the presence of rhamnose [53]. Our goal was to expand these studies to identify functional consequences of rhamnose allosteric signaling in RhaR. We reasoned that rhamnose binding to RhaR might affect the DNA binding affinity of each RhaR protomer in a dimer, and/or the ability of the two DBDs in each RhaR dimer to bind cooperatively to the full RhaR DNA site. Although other mechanisms are certainly possible, based on findings with other AraC family proteins, we hypothesize the following mechanisms. We propose that changes in the DNA-binding affinity of each protomer might involve backbone movement resulting in HTH repositioning, as has been proposed from comparison of AraC family DBD structures in

activating or non-activating conformations [34, 38]. We further propose that changes in the ability of the two DBDs in a RhaR dimer to bind cooperatively to DNA might involve effector-dependent rigid body movement of the DBDs relative to the regulatory domains, similar to the domain movements proposed in the AraC protein light-switch model [35]. EMSAs were used to determine the equilibrium DNA binding affinity of RhaR to its full (dimer binding) and half (protomer binding) DNA binding sites in the presence and absence of L-rhamnose, and the affinity of the RhaR-DBD to the half-site DNA in the absence of L- rhamnose.

To determine whether rhamnose binding affected the DNA binding affinity of each RhaR protomer, we assayed binding of RhaR to a DNA fragment that included a single 17 bp DNA half-site (one protomer binds each half-site) with 4 bp of flanking DNA on either end. This method was used to measure binding by each protomer of dimeric RhaR rather than attempting to assay DNA binding by RhaR monomers based on our finding that there was no detectable RhaR monomer peak following size exclusion chromatography with protein at a concentration of 28 uM (0.9 mg/mL). The short flanking DNA on either end of the half-site sequence in the DNA fragments was expected to preclude binding by both RhaR DBDs of a dimer to a single DNA fragment. Thus, this assay was expected to report the DNA binding affinity of each RhaR protomer of a RhaR dimer. To identify the effect of the RhaR NTD on the function of the RhaR DBD, we also assayed equilibrium DNA binding by the isolated RhaR-DBD to the same half-site DNA fragment. 4 show that RhaR bound 350-fold more tightly to the half-site DNA than the isolated RhaR-DBD, suggesting that the presence of the RhaR NTD in the full-length protein had a large impact on DNA binding affinity. The binding affinity we measured for RhaR binding to half-site DNA was approximately 30-fold tighter than that previously measured by Tobin and Schleif [53], however, they reported fitting a single data point to a Michaelis-Menton

binding curve, and did not report which RhaR half-site was used. We measured half-site DNA binding over RhaR concentrations that yielded 0% (undetectable) to 100% binding of the DNA fragment in both the absence and presence of rhamnose (Fig. 46), suggesting that the current measurements are likely to be more accurate. We were surprised to find that rhamnose did not appreciably alter the binding of RhaR to half-site DNA. Thus, we conclude that the RhaR NTD increased the binding affinity of each RhaR protomer by a mechanism that was independent of rhamnose, and that rhamnose binding caused no further change in protomer binding affinity.

We also assayed the effect of rhamnose on RhaR binding to the full RhaR DNA site (Table 14). RhaR bound to the full DNA site approximately 10-fold more tightly in the presence of rhamnose compared to the absence of rhamnose. This fold difference in binding is consistent with previous findings [53]. Comparison of RhaR binding to half-site DNA with binding to full site DNA showed that binding to the full site was stronger in both the absence and presence of rhamnose. However, examination of the Hill slopes for evidence of cooperative binding (Hill slopes significantly greater than 1), suggested that only the RhaR binding to the full site in the presence of rhamnose exhibited significant cooperativity (data not shown).

Our finding that the strength of RhaR binding to half-site DNA was not rhamnose dependent suggests that the affinity of each protomer for DNA was not increased in the presence of rhamnose. Thus, an alternative mechanism must explain the increase in the strength of RhaR binding to full-site DNA. Based on this, we hypothesized that the cooperative binding in the presence but not the absence of rhamnose might best be explained by a difference in the

Table 14 Apparent K_D values of RhaR and RhaR-DBD to various DNAs

Protein	DNA	Apparent K_D	
		(+)L-rhamnose*	(-)L-rhamnose*
RhaR	Full-Site	4.6×10^{-12} M	4.2×10^{-11} M
	Gapped	1.3×10^{-10} M	2.7×10^{-10} M
	Half-Site	2.5×10^{-9} M	3.1×10^{-9} M
RhaR-DBD	Half-Site	N.A[a]	1.1×10^{-6} M

* Standard errors were no greater than 11% of the value.

[a] The Kd measurement for RhaR-DBD binding to half-site DNA was measured only in the absence of L-rhamnose.

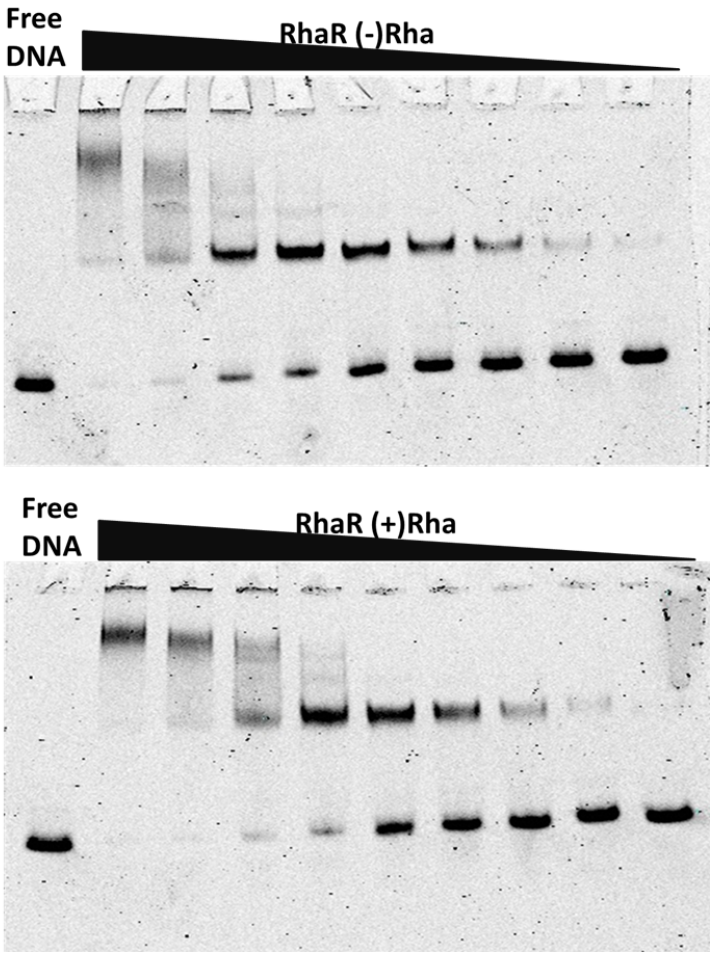


Fig. 46. Representative EMSA gels showing the binding of RhaR-His6 to half site DNA. Top, RhaR-His6 binding to half site DNA in the absence of L-rhamnose [(-)rha]. Bottom, RhaR-His6 binding to half site DNA in the presence of L-rhamonse [(+)rha]. Free DNA (F) is in the first lane on each gel. The black triangles represent decreasing concentrations of proteins, with serial two-fold dilutions. The highest protein concentration added was at 6×10^{-8} M.

positioning or the flexibility/dynamics of the positioning of the two DBDs in the dimer. More specifically, we hypothesized that in the absence of rhamnose, the two DBDs in a RhaR dimer were somewhat constrained such that their binding to the full RhaR binding site was not significantly cooperative; but in the presence of rhamnose, the two DBDs were either more optimally positioned, or more flexible/dynamic in their positioning, such that binding to the full site DNA was cooperative.

To test the hypothesis that rhamnose binding to RhaR might alter the ability of the DBDs in a RhaR dimer to bind cooperatively to full-site DNA, we tested RhaR binding to "gapped" full-site DNA. This DNA fragment had double-stranded half-sites but had 17 bp of single stranded DNA between the half-sites – a region of DNA that is not contacted by RhaR. The rationale, as previously proposed by Harmer *et. al* [29], was that the single stranded spacer DNA would provide much greater flexibility than the double stranded spacer, and thus would be better able to accommodate binding to two RhaR DBDs that might be sub-optimally positioned in the absence of rhamnose to bind to the fully double stranded full site DNA. RhaR binding to the gapped DNA differed by only two-fold in the absence versus the presence of rhamnose, which is a significantly smaller difference than the 10-fold difference measured for fully double stranded full site DNA. This finding supports the hypothesis that at least the majority of the effect of rhamnose binding to RhaR is a change in the relative positioning or the flexibility/dynamics of the two DBDs in the dimer such that they are better able to bind cooperatively to the full RhaR DNA binding site.

RhaR DNA bending. The hypothesis that the primary consequence of rhamnose binding to dimeric RhaR may involve the repositioning of the two DBDs relative to one another suggests

that there might be a measurable rhamnose-dependent difference in the degree of DNA bending by RhaR when bound to its full DNA site. Tobin and Schleif previously estimated the degree of RhaR bending in the presence of rhamnose to be approximately 160°, but did not measure bending in the absence of rhamnose [53]. We used the gel electrophoresis assay developed by Wu & Crothers [103] to estimate the bending angle of RhaR binding to full site DNA in the absence and presence of rhamnose. We designed five 292 bp DNA fragments with the 50 bp RhaR DNA binding site at different positions (Fig.47A), and otherwise identical DNA sequences. Figure 47B shows that the equivalent length DNA fragments had a faster mobility when the RhaR-bound DNA binding sequence was located at one or the other end of the DNA fragment than when the site was located at or nearer to the center of the fragment. The gel also shows that there were differences in the migration of the identical RhaR-bound DNA fragments in the absence versus the presence of rhamnose. Using the method described by Thompson & Landy [104], we estimated the bending angle with RhaR bound to its full DNA site to be 141° in the absence of rhamnose and 150° in the presence of rhamnose, comparable to the prior measurement in the presence of rhamnose. The large difference in the length of the DNA fragments used for the bending experiments (Tobin, 787 bp; this work, 292 bp) may account for at least some of the difference in the bending estimates in the presence of rhamnose between these two studies. The difference between the bend angles in our experiments in the absence and presence of rhamnose was found to be statistically significant, with a P value of <0.005. Although this difference in bend angles was modest, this is consistent with the relatively small (~10x) difference in binding affinity of RhaR for its full DNA site in the absence and presence of rhamnose. Interestingly, despite the AraC protein's switch from looping in the absence of arabinose to binding adjacent DNA sites in the presence of arabinose, and the use of a phase-

sensitive DNA bending assay, no arabinose dependent change in the angle of DNA bending by AraC has been detected [155].

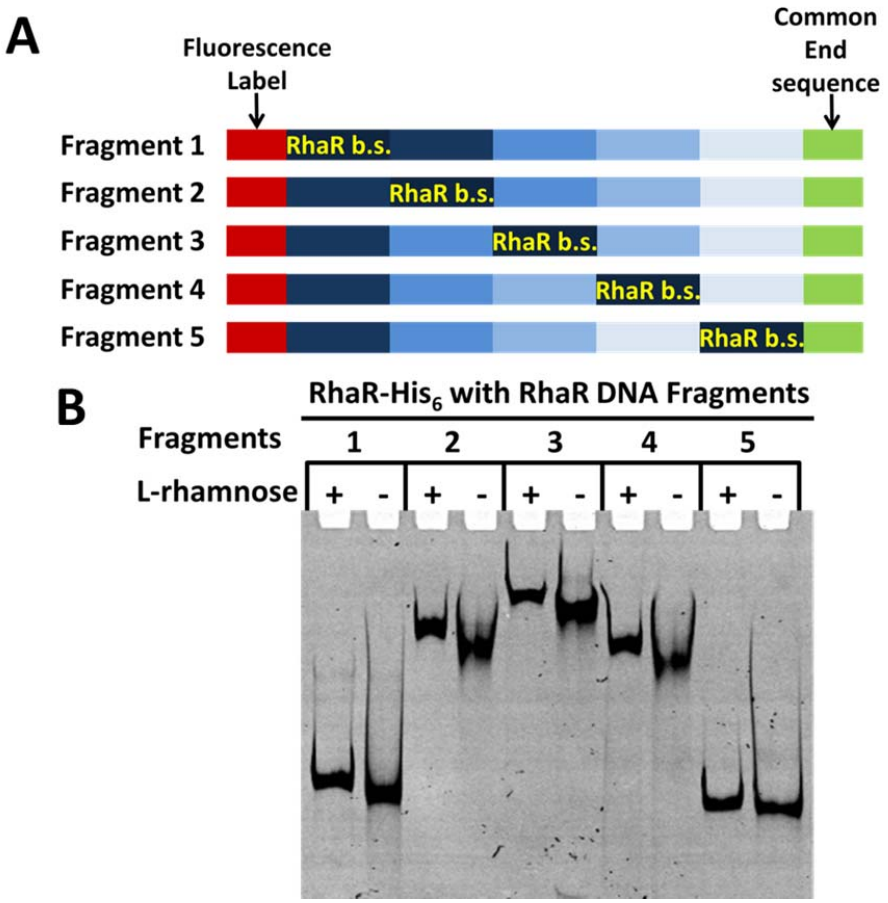


Fig. 47. RhaR DNA bending as measured by EMSA. (A) Maps of DNA fragments to estimate DNA bending by RhaR. Each DNA fragment was labeled with a fluorescent tag at the 5'. RhaR binding site positions were included, and different shades of blue blocks represent the same length but different sequences of the flank region. (B) EMSA showing RhaR-His_6 protein binding to different DNA fragments for bending in the presence (+) and absence (-) of L-rhamnose.

Chapter 9

DNA binding and Allosteric signaling by RhaS

The AraC family transcriptional activator RhaS regulates the expression of genes involved in the L-rhamnose regulon. RhaS activates transcription of both the *rhaBAD* operon, which contains genes encoding L-rhamnose catabolic enzymes, and the *rhaT* operon, which contains a gene encoding L-rhamnose transporter [10, 49]. RhaS also negatively auto-regulates its own expression, and the expression of RhaR, at the *rhaSR* operon. In the presence of rhamnose, RhaS activates transcription approximately 1000-fold higher than in the absence of rhamnose [51]. This increase in transcription activation suggests that there must be an allosteric signal that is transmitted from the regulatory domain where the rhamnose binds to the DNA-binding domain between its (-)rha and (+)rha states. Although RhaS DNA-binding affinity isn't the only property that is affected by L-rhamnose binding, measurements of L-rhamnose-dependent effects on DNA binding by RhaS are expected to provide information about the mechanisms of allosteric L-rhamnose signalling in RhaR. Here, by using EMSA assays, I have measured the dissociation constants for the interactions between RhaS and the isolated RhaS-DBD with RhaS DNA binding sites.

Purification of RhaS-His_6 in the presence of rhamnose. To overproduce soluble RhaS protein, 0.2% rhamnose (final concentration) was added before induction. RhaS protein was first purified as C-terminal Hi_6 fusion protein by Ni^{2+} affinity chromatography. During the purification, 0.2% rhamnose was also included in the purification buffer (binding, wash, elution buffers). After eluting RhaS-His_6 protein from the Ni^{2+} column, the protein sample was quickly loaded onto a a HiPrep 16/60 Sephacryl S-300 HR column (GE healthcare) and eluted with buffer without the rhamnose. By using this method, I was able to purify 4 mg of RhaS-His_6 protein from 1L of cultured bacterial cells (Fig. 48A). The purified RhaS was active, based on the evidence that it bound to the DNA in an EMSA, as shown in Fig. 48B.

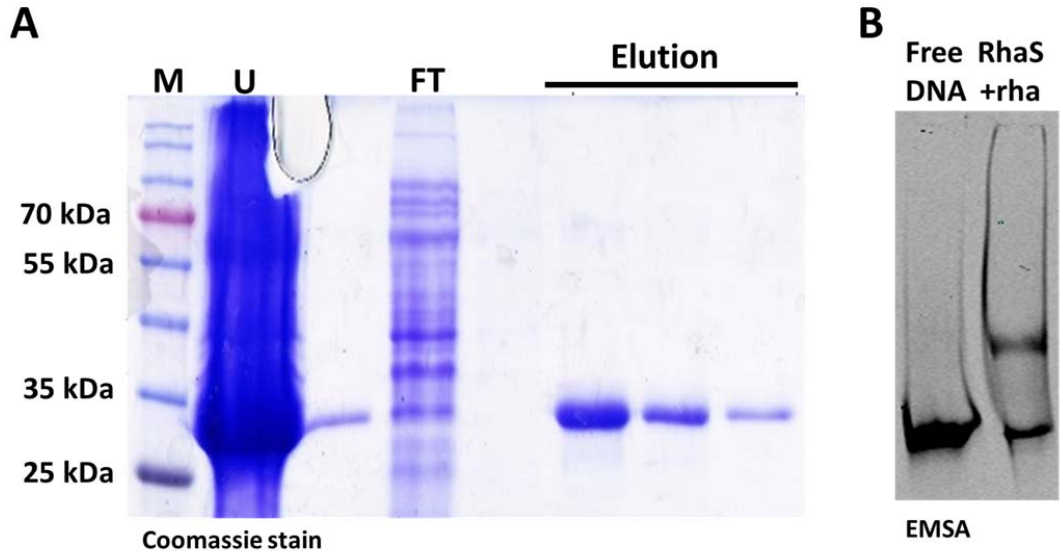


Fig. 48. (A) 12% SDS-PAGE gel stained with coomassie blue showing purified RhaS-His_6 protein eluted from Ni^{2+}-affinity chromatography. Molecular mass marker (M) is in the first lane. (B) The EMSA gel showing the DNA-binding by purified RhaS with L-rhamnose (+rha).

Equilibrium binding measurements. EMSAs were used to determine the equilibrium DNA binding affinity of RhaS to its full (dimer binding) and half (protomer binding) DNA binding sites in the presence of rhamnose. Because we were unable to purify RhaS in the absence of rhamnose due to insolubility of the protein, the equilibrium binding assays for the full length RhaS were carried out in the presence of rhamnose. To determine the effect of rhamnose binding on the DNA binding affinity of each RhaS protomer, I have assayed full length RhaS binding to a DNA fragment that included a single 17 bp DNA half-site (one protomer binds each half-site) with 4 bp of flanking DNA on either end. Because the short flanking DNA on either end of the half-site sequence in the DNA fragments was expected to preclude binding by both RhaS DBDs of a dimer to a single DNA fragment, this assay was expected to report the DNA binding affinity of each RhaS protomer of a RhaS dimer. To identify the effect of the RhaS NTD on the function of the RhaS DBD, the affinity of the RhaS-DBD to the same half-site DNA in the absence of L- rhamnose was also measured.

The results in Table 15 show that in the presence of rhamnose full length RhaS bound 40-fold more weakly to the half-site DNA than the isolated RhaS-DBD, suggesting that the presence of the RhaS NTD in the full-length protein decreased the RhaS-DBD DNA binding affinity to the half site in the presence of rhamnose. We also found that RhaS bound to the full DNA site approximately 10-fold more tightly than to the half site in the presence of rhamnose, suggesting that the two DBDs in the RhaS dimer bind cooperatively. The finding that the full length RhaS binding to the full-site DNA was about 3-fold weaker than the isolated RhaS-DBD binding to the half-site indicated that despite the increased cooperativity in full site RhaS DNA binding, the RhaS-NTD decreased DNA binding. Although we have not measured the binding affinity of the

full length RhaS to the full site in the absence of rhamnose, our *in vivo* transcription activation assay predicted that it will be approximately 1000-fold weaker relative to (+)rha.

RhaS DNA bending. Since we do not have RhaS protein purified in the absence of rhamnose, we can only measure the DNA bending by RhaS in the present of rhamnose. I have designed five 292 bp DNA fragments with the 51 bp RhaS DNA binding site at different positions. These DNA fragments are similar to those used in the RhaR bending assay. The flanking sequences are identical DNA sequences. Using gel electrophoresis assays and the method described by Thompson & Landy [104], we estimated the bending angle with RhaS bound to its full DNA site to be 105° in the absence of rhamnose (Fig. 49A). Each DNA-protein band on the gel smeared during gel running, which made the measuring quite difficult. In order to confirm the bend angle induced by RhaS, I have purified a variant of RhaS protein, $RhaS^{L201R}$-His_6 [52] which binds DNA more tightly than the wild-type RhaS protein, and performed bending assay with this variant protein. I found that $RhaS^{L201R}$-His_6 also produced an estimated 105° bend when binding to the DNA (Fig. 49B), which was consistent with the bend angle produced by the wild-type RhaS protein. The RhaR DNA contains a natural bend with an estimated 58° bend angle in the absence of RhaR protein bound (Fig. 50A). The RhaS DNA, on the other hand, does not contain a natural bend (Fig. 50B).

Prior data indicated that RhaS can bind to the RhaR binding site, but with decreased *rhaSR* expression [156]. Therefore, we previously hypothesized that RhaS decreases *rhaSR* expression by not bending the DNA sufficiently for αCTD to contact CRP. To test this hypothesis, we measured the DNA bending by RhaS at the RhaR DNA binding site. We found that RhaS produced a 148° bend (Fig. 51). This bend angle was 43° greater than the bend angle RhaS produced at the RhaS binding site, supporting the hypothesis for strong auto-regulation.

However, in our assays, RhaR bent RhaR binding site to 150° in the presence of rhamnose, which was approximately the same as the bend angle produced by RhaS at the RhaR binding site (within the error of our measurements). This result indicated that DNA bending cannot explain RhaS auto-regulation.

Table 15 Apparent K_D values of RhaR and RhaR-DBD to various DNAs

Protein	DNA	Apparent K_D	
		(+)L-rhamnose	(-)L-rhamnose
RhaS	Full-Site	3.8×10^{-9} M	N.A.[b]
	Half-Site	4.5×10^{-8} M	N.A.[b]
RhaS-DBD	Half-Site	N.A.[a]	1.2×10^{-9} M

[a] The K_d measurement for RhaS-DBD binding to half-site DNA was measured only in the absence of L-rhamnose.

[b] The K_d measurement for RhaS binding to full-site and half-site DNA were measured only in the presence of L-rhamnose.

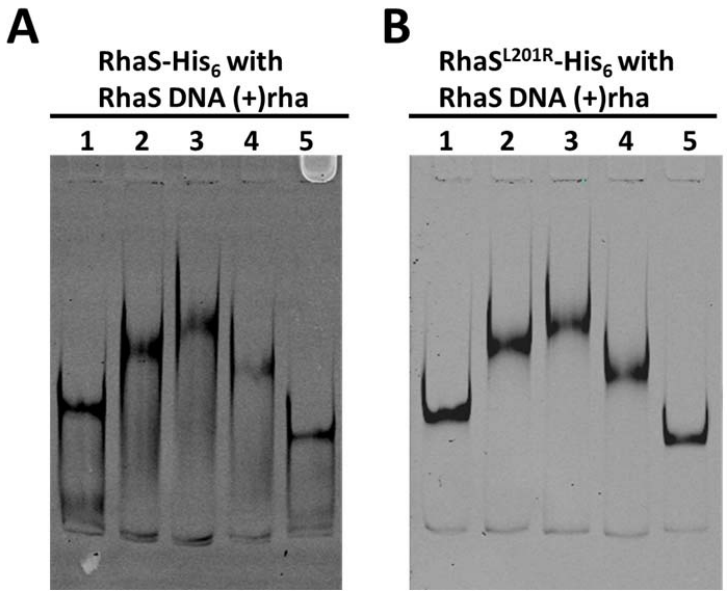


Fig. 49. EMSA showing **(A)** wild-type RhaS-His_6 and **(B)** $RhaS^{L201R}$-His_6 binding to different DNA fragments, and both produced a105° bend in the presence of L-rhamnose [(+)rha].

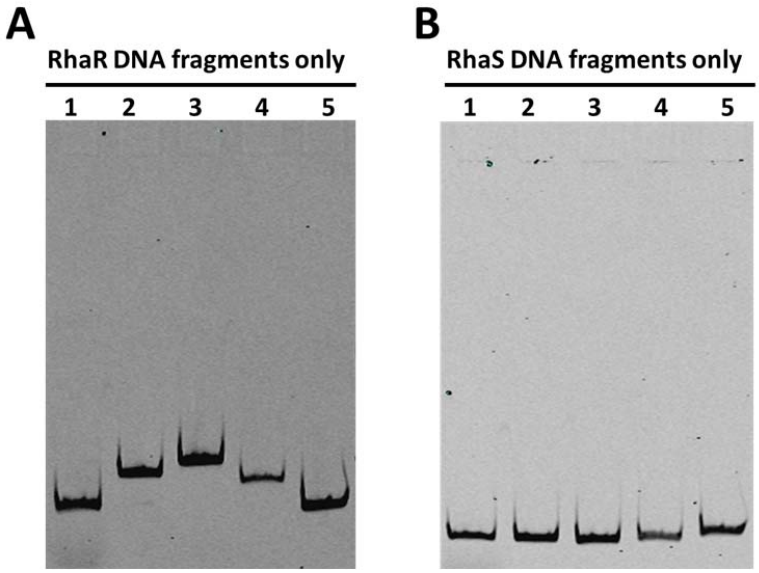


Fig. 50. EMSA showing **(A)** RhaR DNA fragments produced a 58° natural bend in the absence of protein, and **(B)** RhaS DNA fragments did not produce bend in the absence of proteins.

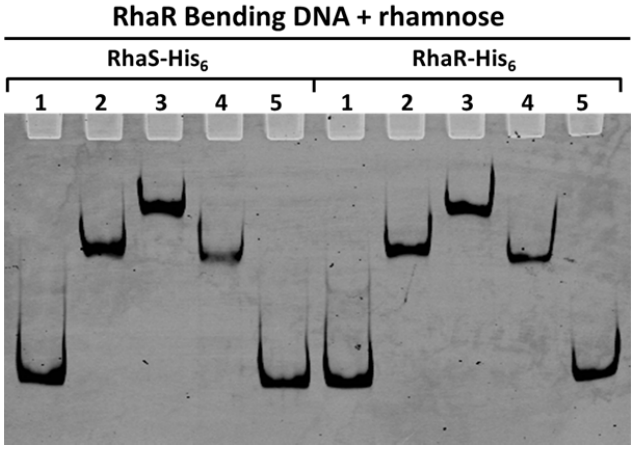


Fig. 51. EMSA showing RhaS-His_6 and RhaR-His_6 bind to the RhaR DNA fragments, producing a 148°bend and 150 °bend, respectively

Chapter 10

Summary of Mutagenesis study of RhaS

I have investigated of residues in the regions of RhaS DBD that were predicted to contact the NTD in allosteric signalling based on alignment with AraC: HTH1, HTH2, and allosteric site in subdomain 2 (AS2). In total, 18 positions within these three regions of RhaS DBD were subjected to site-directed random mutagenesis, and assayed their transcription activation activities in the *rhaB-lacZ* fusion. Table 16, 17, and 18 summarized the transcription activation activates and protein levels of all the RhaS variants that I constructed.

Table 16. Transcription Activation of *rhaB-lacZ* and RhaS AS2 Variants[a].

Substitution	(-)rha		(+)rha	
	% WT Activation[a]	% WT Protein Levels[b]	% WT Activation[a]	% WT Protein Levels[b]
WT	100 ± 6	100 ± 3	100 ± 5	100 ± 2
L208M	1204 ± 20	97 ± 3	126 ± 6	98 ± 3
L208I	313 ± 8	95 ± 3	119 ± 4	99 ± 2
Q210R	83 ± 6	102 ± 1	103 ± 10	104 ± 5
Q210S	115 ± 11	98 ± 2	110 ± 12	103 ± 1
Q210G	42 ± 4	86 ± 3	103 ± 13	88 ± 4
Q210N	72 ± 7	97 ± 7	96 ± 10	92 ± 4
Q210K	875 ± 17	98 ± 8	132 ± 14	97 ± 1
Q211G	81 ± 11	91 ± 8	139 ± 13	87 ± 5
Q211V	120 ± 12	95 ± 6	153 ± 13	86 ± 2
Q211I	92 ± 13	62 ± 4	58 ± 6	68 ± 4
T212C	19 ± 3	85 ± 3	71 ± 5	89 ± 6
T212M	9 ± 2	89 ± 9	9 ± 0.4	97 ± 9
T212A	8 ± 1	71 ± 6	45 ± 2	94 ± 2
L214V	16 ± 2	114 ± 5	90 ± 8	86 ± 4
L214I	41 ± 8	103 ± 5	136 ± 14	94 ± 3
L214M	70 ± 6	106 ± 9	86 ± 8	86 ± 8
L214G	4 ± 1	95 ± 9	19 ± 1	92 ± 5
R218V	12 ± 2	99 ± 2	90 ± 7	88 ± 5
R218L	6 ± 1	79 ± 7	11 ± 1	94 ± 2
R218G	7 ± 1	87 ± 3	32 ± 2	81 ± 4
Y219F	20 ± 2	96 ± 5	115 ± 7	82 ± 3
Y219L	9 ± 1	84 ± 3	36 ± 3	88 ± 6
Y219M	6 ± 1	94 ±9	12 ± 1	92 ± 4

[a.] Variants were assayed in groups with a wild type RhaS activity range of 0.3-0.4 Miller Units (-)rha and 208-217 Miller Units (+)rha. Values are the average of at least two independent assays with two replicates each and are shown as the percent of the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

[b.] Quantification of variant protein levels compared to wild-type RhaS. Total protein was separated by 12% SDS-PAGE. Proteins were transferred to nitrocellulose, probed with anti-RhaS and anti-DnaK antibodies and quantified using densitometry analysis. RhaS values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values.

Table 17. Transcription Activation of *rhaB-lacZ* and RhaS HTH1 Variants[a].

Substitution	(-)rha		(+)rha	
	% WT Activation[a]	% WT Protein Levels[b]	% WT Activation[a]	% WT Protein Levels[b]
WT	100 ± 6	100 ± 3	100 ± 5	100 ± 2
R229I	69 ± 5	79 ± 6	58 ± 1	100 ± 7
R229G	14 ± 1	79 ± 7	3 ±0.2	88 ± 4
R229D	20 ± 1	74 ± 8	7 ± 0.3	94 ± 3
R229T	29 ± 2	72 ± 6	22 ± 1	66 ± 1
R229P	28 ± 2	72 ± 3	2 ± 0.2	65 ± 3
H230E	393 ± 27	93 ± 7	67 ± 2	90 ± 6
H230V	31 ± 2	70 ± 2	22 ± 1	91 ± 3
H230I	27 ± 3	114 ± 1	3 ± 0.5	120 ± 6
H230Q	69 ± 6	78 ± 4	57 ± 2	76 ± 4
H230L	25 ± 2	66 ± 6	14 ± 0.5	68 ± 5
H230G	20 ± 1	65 ± 3	21 ± 1	77 ± 3
H230R	21 ± 2	73 ± 6	11 ± 0.5	67 ± 9
L232M	9 ± 1	70 ± 5	11 ± 0.4	63 ± 3
L232V	16 ± 2	66 ± 3	40 ± 1	59 ± 2
L232I	32 ± 3	71 ± 8	76 ± 2	60 ± 4
R233L	72 ± 3	99 ± 1	25 ± 1	103 ± 2
R233P	8 ± 1	85 ± 8	1 ± 0.1	85 ± 6
R233A	23 ± 1	80 ± 3	15 ± 1	89 ± 6
R233W	13 ± 0.3	82 ± 6	7 ±0.2	93 ± 7
R233D	12 ± 1	79 ± 2	5 ± 0.2	94 ± 3
R233G	14 ± 1	74 ± 5	7 ± 0.2	92 ± 3
H234T	295 ± 15	86 ± 6	131 ± 7	85 ± 4
H234I	70 ± 11	102 ± 9	123 ± 7	93 ± 4
H234K	65 ± 16	90 ± 7	106 ± 5	91 ± 7
H234G	39 ± 5	91 ± 8	49 ± 3	92 ± 6
H234Y	83 ± 4	89 ± 7	94 ± 6	84 ± 6
H234V	72 ± 9	84 ± 5	104 ± 6	80 ± 4
H234D	30 ± 1	94 ± 1	42 ± 2	85 ± 4
H234E	57 ± 4	97 ± 3	56 ± 3	88 ± 1

[a.] Variants were assayed in groups with a wild type RhaS activity range of 0.3-0.4 Miller Units (-)rha and 208-217 Miller Units (+)rha. Values are the average of at least two independent assays with two replicates each and are shown as the percent of the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

[b.] Quantification of variant protein levels compared to wild-type RhaS. Total protein was separated by 12% SDS-PAGE. Proteins were transferred to nitrocellulose, probed with anti-

RhaS and anti-DnaK antibodies and quantified using densitometry analysis. RhaS values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values.

Table 18. Transcription Activation of *rhaB-lacZ* and RhaS HTH2 Variants[a].

Substitution	(-)rha		(+)rha	
	% WT Activation[a]	% WT Protein Levels[b]	% WT Activation[a]	% WT Protein Levels[b]
WT	100 ± 6	100 ± 3	100 ± 5	100 ± 2
E261N	74 ± 5	102 ± 1	101 ± 6	97 ± 3
E261G	15 ± 1	91 ± 8	11 ± 1	89 ± 8
E261S	1163 ± 92	128 ± 7	229 ± 7	127 ± 6
E261M	482 ± 39	92 ± 3	176 ± 6	118 ± 3
E261F	640 ± 56	95 ± 8	224 ± 4	107 ± 6
F262Y	48 ± 4	95 ± 4	100 ± 5	103 ±8
F262L	11 ± 1	96 ± 2	3 ± 1	92 ± 8
N263Y	22 ± 1	100 ± 8	45 ± 3	104 ± 6
N263E	33 ± 2	108 ± 7	72 ± 4	92 ± 1
N263R	46 ± 4	100 ± 5	85 ± 3	99 ± 5
N263L	19 ± 1	128 ± 8	13 ± 2	99 ± 4
N263W	14 ± 1	93 ± 2	26 ± 3	96 ± 7
W264H	95 ± 6	90 ± 5	77 ± 4	105 ± 7
W264M	377 ± 42	80 ± 5	69 ± 4	93 ± 7
W264E	56 ± 2	90 ± 5	28 ± 2	100 ± 5

[a.] Variants were assayed in groups with a wild type RhaS activity range of 0.3-0.4 Miller Units (-)rha and 208-217 Miller Units (+)rha. Values are the average of at least two independent assays with two replicates each and are shown as the percent of the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

[b.] Quantification of variant protein levels compared to wild-type RhaS. Total protein was separated by 12% SDS-PAGE. Proteins were transferred to nitrocellulose, probed with anti-RhaS and anti-DnaK antibodies and quantified using densitometry analysis. RhaS values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-)rha or (+)rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values.

Chapter 11

Discussion

Identifying Small Molecule Inhibitor Binding Site in RhaS. In this study, we have demonstrated that SE-1 directly bound to VirF and RhaS-DBD using florescence-based thermal melt assay and intrinsic tryptophan florescence assay. Mutagenesis studies suggested that residues D191, S249, and L257 were at or near the SE-1 binding site on RhaS-DBD, which supports the prediction of docking that SE-1 is likely bound to a small pocket between the two HTH motifs, and these residues also contribute to the binding of SE-1. Identification of the binding site of SE-1 in RhaS allows us to postulate the possible mechanism of action of SE-1. It is likely that upon binding to the small pocket between the two HTH motifs, SE-1 contacts residues D191, S249, and L257. Because of the binding of SE-1, RhaS is predicted to be unable to contact DNA, and therefore, will be unable to activate transcription of the *rhaBAD* and *rhaT* operons. AraC family proteins are defined by the presence of a structurally conserved DBD. A small molecule inhibitor targeting this domain could potentially target many protein members of the family. To target multiple members of the AraC family, the small molecule must bind to a pocket with a conserved shape to accommodate the shape of the DNA across family members. Additionally, sequence alignment and analysis of 200 AraC family members indicate that residues which possibly contribute to SE-1 binding are fairly conserved. Position D191, S239, L247 shared 50%, 36%, and 53% sequence similarity with the 200 AraC members analyzed. In our lab, we have found that SE-1 inhibited the activities of multiple AraC family members: RhaS, RhaR, VirF, Rns, and ToxT. Our docking studies indicated that SE-1 could fit in the small pocket between the two HTH motifs in all the proteins tested, including MarA, RhaS, VirF, and Rns. Therefore, it is likely that SE-1 inhibits these proteins by binding to the same pocket between the HTH motifs. Since the HTH motifs are fairly well conserved across the AraC family, and docking studies suggest that the binding pocket that SE-1 binds to is present in all

AraC family proteins tested, it is possible that SE-1 could target quite a large number of AraC family proteins including those regulate expression of virulence genes. Traditional antibiotics act by inhibiting bacterial functions (such as cell wall synthesis, DNA replication, RNA transcription and protein synthesis) that are essential for cell growth. Targeting the expression of virulence factors does not affect the cellular viability. It has several advantages including exerting less selective pressure, which may result in decreased resistance. Therefore, small molecule inhibitor SE-1 that target AraC family virulence activators potentially can be developed into a "broad spectrum" antimicrobial agent.

Higher resolution structure of ToxT and SE-1-ToxT crystallization trial. In an effort to obtain crystals of ToxT in a complex with the inhibitor SE-1 with the goal of more directly identifying the SE-1 binding site, we first started with crystalizing ToxT. Despite following the protocol used by Lowden *et al*. as closely as possible, we were unable to obtain ToxT crystals using the published crystallization conditions [38]. This wasn't entirely surprising giving that poor reproducibility has been a notorious problem in the protein crystallization field [157, 158], and it is common that crystallization results differ even under the "same conditions" [159]. Many parameters, such as the precise protein concentration, temperature, buffer composition and pH, among other factors, can affect protein crystallization [160, 161]. After screening over 480 different crystallization conditions, we obtained crystals of ToxT with a condition that is completely different from the published one. The structure was solved to 1.65Å resolution (PDB code 4MLO) which is a higher resolution than the previously determined ToxT structure. Many structural quality parameters of the current ToxT structure (R_{free}, clashscore, etc.) were in the high percentile range, suggesting the overall high quality of the structure, and the higher quality of the structure than the previous structure (Fig. 52).

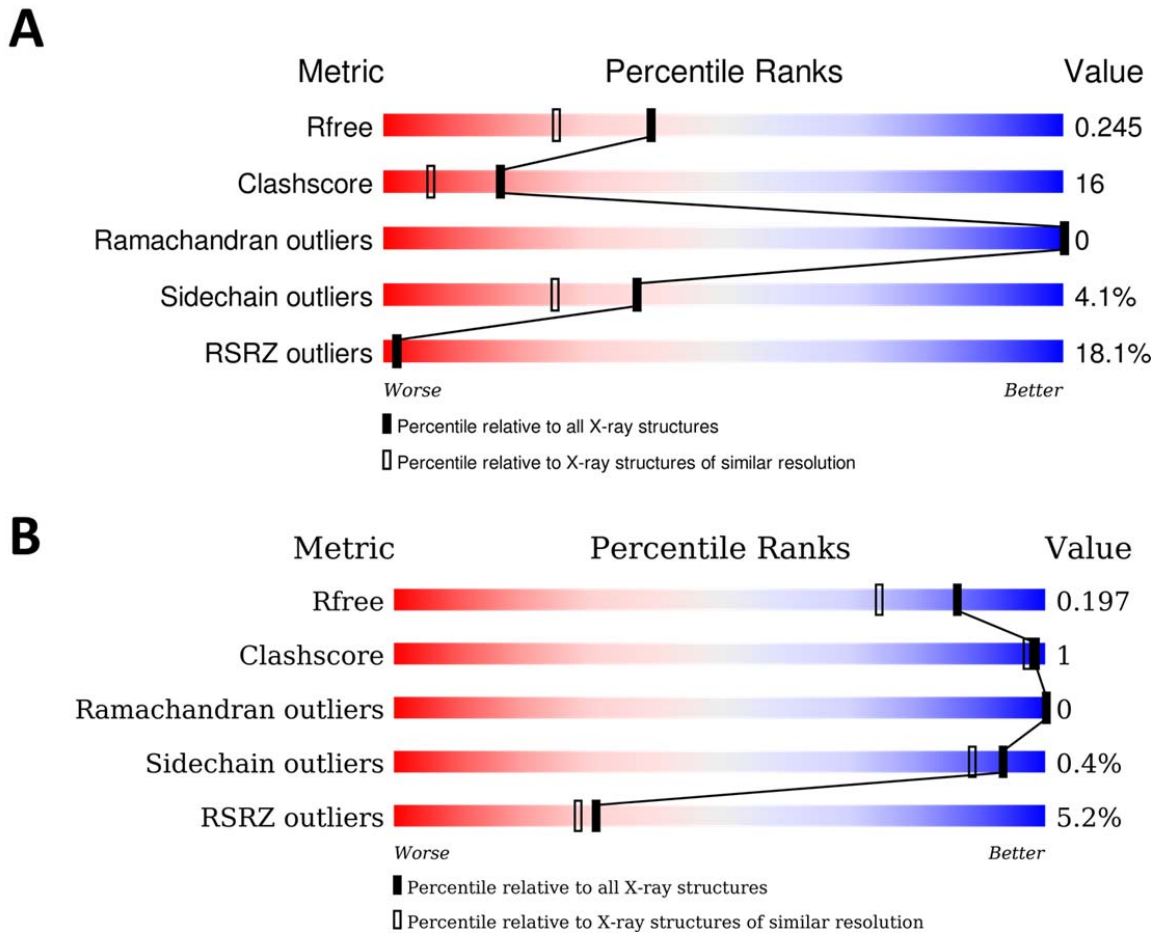


Fig. 52. Percentile scores (ranging between 0-100) for global validation metrics of **(A)** previously determined ToxT structure (PDB:3GBG), and **(B)** current ToxT structure (PDB:4MLO). From [162] and [163].

In addition to the improved overall quality of the current structure, a previously disordered region of ToxT between α1 and β9 from residues D101 to E110, could be fully traced in the current structure. Many residues in this region have previously been shown important for ToxT activity in respond to its fatty acid effectors and Virstatin, an inhibitor of ToxT activity [41, 79, 164], thus solving the structure of this region provided significant new information for understanding the role of individual residues within and around this region. For instance, alanine substitutions of residues M103, R105 and N106 within this region have shown activity at the *ctxA* promotor that is three-fold higher than wild-type ToxT, suggesting that these residues have a negative effect on the activity of wild-type ToxT [164]. The analysis of the current ToxT structure reveals that R105 forms a salt bridge with E156. This observation combined with the alanine substitution results suggests the possibility that R105 holds E156 in a position that somewhat attenuates ToxT activity. E156 is located in helix α3, which is likely involved in dimerization to facilitate transcriptional activation [38]. Thus, R105 may maintain the activity of ToxT at its wild-type level by somewhat supressing dimerization (relative to the R105A substitution). Of course, protein structure does not always provide explanations for the phenotypes of protein variants. For example, other than their potential effects on R105, the structure does not provide obvious explanations for how alanine substitutions at residues M103 or N106 also increased ToxT activity. Virstatin, a small molecule inhibitor of ToxT identified by Hung *et al* (2005) [79], blocks ToxT dimerization and thus its ability to activate transcription of the *tcp* and *ctx* promoters [165]. Shaknovich *et al*. [165] also have demonstrated that a ToxT variant, L114P, is resistant to virstatin and suggested that L114P may result in a conformational change in ToxT that allows the protein to dimerize more efficiently [165]. Lowden *et al*. [38] suggested that the previously disordered region from D101 to E110 might be involved in the

virstatin resistance of the L114P variant due to their proximity; however, there are no obvious interactions between L114 and any of the residues in the 101-110 region that would suggest that region's involvement in the mechanism of virstatin resistance of ToxT L114P.

Once we obtain the ToxT crystal, we attempted to crystalize ToxT in a complex with the inhibitor SE-1. However, obtaining SE-1-ToxT crystals either by soaking or direct crystallization was not successful. One explanation for failing to obtain SE-1-ToxT crystals was that SE-1 bound to ToxT with a much weaker affinity than the other proteins we identified as inhibited by SE-1. Therefore, it might require ToxT crystals to be soaked in higher concentration of SE-1 than those used, or increased SE-1 concentration while mixing with ToxT protein, but, the limited solubility of SE-1 made it impossible to test these possibilities. Another possibility is that the current ToxT structure contains the effector PMA, and thus the protein was in its non-activating state which is unable to bind to DNA [38]. Since the switch from the activating to the non-activating state of ToxT is predicted to involve a shift in the positioning of the two HTH motifs relative to one another, it's possible that the binding pocket that SE-1 binds to is not present in the non-activating state of ToxT [34, 38]. However, it's likely that at least some portion of the protein used to set up crystallization trays was in an activating state initially, as we showed that adding oleic acid, which is predicted to act as a ToxT effector in the same way that PMA does, inhibited DNA binding. The finding that SE-1 inhibited DNA binding by ToxT suggested that SE-1 can at least bind to the activating conformation of ToxT, but we don't know whether it can bind to the non-activating conformation. Therefore, one hypothesis is that SE-1 can only bind to the activating conformation of ToxT, and we either haven't been able to find a condition to obtain crystals of the activating ToxT conformation, or that the activating conformation of ToxT has too much inherent flexibility to readily form crystals. Alternatively,

SE-1 might be able to bind both conformations of ToxT, but with weak affinity, thus it cannot easily be crystalized with any forms of ToxT.

SE-1 optimization. Typically, in the process of drug discovery, small molecule hits from a high throughput screen require several rounds of optimization to improve the specificity and potency before being considered lead compounds that have the potential to enter the clinical drug development process [166-169]. Hence, in collaboration with Dr. Jeff Aubé, we have carried out optimization studies on the small molecule inhibitor SE-1 to improve its potency and specificity. Unfortunately, none of the analogues tested had increased inhibition potency or specificity. Our results indicated that the biquinolinium core of SE-1 is important for the inhibition activity against the AraC family activator VirF. Replacing the quinolinium core resulted in inactivity of the analogues (entries 6, 7, 12, 13, 15, Table 9). To maintain activity, the N-substitution of the *n*-butyl group had to be limited to linear alkyls (entries 2, 19-23, Table 9); and any bulky groups (i.e. chlorobenzyl group) resulted in analogues that were less active than SE-1 (entry 3 & 5, Table 9). However, the *n*-nonyl substitution showed loss of selectively (entry 22, Table 9), while the *n*-hexyl substituted analogue showed similar inhibition activity and specificity as compared to SE-1 (entry 21, Table 9). The above findings suggested that only short length of the alkyls (maybe up to six-carbon alkyl) can be tolerated at this position to maintain the selectivity of the analogues, although it was unclear that how *n*-heptyl, and *n*-octyl would behave in the inhibition assay. The reason for this phenomenon was not known.

During the optimization, it was discovered that SE-1 and several analogues converted to quinolinium salts when dissolved in aqueous solution, and further, that the quinolinium salt was responsible for the observed inhibitory activity. The quinolinium salt is positively charged, and can accept an electron pair in order to bond to a nucleophile [133]. When this reaction occurs in

cell, it could cause undesired disposition (including absorption, distribution, metabolism, and excretion), and even cell damage [170-172], therefore, quaternary ammonium salts such as this have limited pharmaceutical relevance. Based on this finding, the Aubé lab was no longer interested in continuing our collaboration. However, our previous study of SE-1 has shown that SE-1 had no detectable effects on the metabolic activity of mammalian tissue culture cells (L-929 cells) at concentrations as high as 40 μM [80]. This suggested that SE-1 might still have the potential to be developed into novel antimicrobial agent despite being a quinolinium salt.

In the current optimization study, we explored the SAR of SE-1 with the goals of improving the potency and specificity of SE-1. However, efforts were not directed specifically at the goal of optimizing the solubility of SE-1. Our preliminary study shows that SE-1 is quite hydrophobic and dissolves poorly in aqueous solution or buffer. Improving the solubility of SE-1 could allow us to further characterize the binding of SE-1 to AraC family proteins. For instance, we might be able to measure the binding affinity of SE-1 for purified VirF protein using fluorescence-based thermal melting assays, for which concentrations of SE-1 that are higher than those currently possible are required. Therefore, optimization of the solubility of SE-1 might enable further studies.

RhaR purification and annotation error. As is the case with many of the AraC family transcriptional activators, many aspects of RhaR function have been difficult to characterize due to the protein's low solubility. During our efforts to optimize RhaR purification, we discovered that the true start site of *rhaR* was a GTG codon located 30 codons downstream of the previously annotated, in-frame +30*rhaR* ATG start codon. Despite cloning the incorrectly annotated open reading frame, prior published purifications yielded only the correct, 282 residue native RhaR protein, likely with the following explanations. Tobin and Schleif [53] used DNA affinity

chromatography in their RhaR purification, thereby selecting for the active 282 residue species. In addition, the Shine-Dalgarno sequence in their clone was not as strong as that in newer vectors such as pET24b, and consequently was unlikely to drive expression of much +30RhaR. Our previously published *in vitro* RhaR studies used protein expressed with an intein-chitin binding domain fused at the C-terminus of RhaR [173]. In retrospect, it appears that RhaR and +30RhaR bound to the chitin column, but that the intein self-cleavage of the RhaR fusion was much more efficient than for the +30RhaR fusion (perhaps due to the higher propensity of +30RhaR to aggregate), again resulting in the purification of predominantly the native RhaR species. Thus, the selection of methods that yielded active protein enabled studies of the native RhaR protein prior to recognition of the incorrect annotation. After identifying the correct *rhaR* translation start site, it was found that both the yield and the solubility of RhaR were dramatically increased.

Structures of RhaR-NTD. After many years of effort, we have finally obtained a crystal structure for a domain of the RhaS or RhaR proteins. The structures of rhamnose-bound [(+) rha] and rhamnose-free [(-) rha] RhaR-NTD were determined at 2.05 Å and 1.73 Å, respectively.

The structures of RhaR-NTD reveal that a metal ion was present deep within the rhamnose-binding pocket, and using X-ray fluorescence and ICP-OES, we determined that the metal was Ni^{2+}. Given that the protein was purified by Ni^{2+}-affinity chromatography, it is likely that the Ni^{2+} displaced the native metal or there was no metal bound to the purified RhaR-NTD, and that Ni^{2+} is not likely the native metal in RhaR. Although metal binding is very common among cupin superfamily proteins [138, 150], among AraC family proteins with a cupin superfamily regulatory domain, RhaR-NTD is the only one that has thusfar been demonstrated to bind a metal ion within the effector-binding pocket. Typical cupin domains contain two conserved motifs with the first motif contributing two His residues and one Glu residue and the

second motif contributing the third His residue to the overall metal binding motif [150]. As in RhaR-NTD, these four residues coordinate a metal ion, that in enzymatic cupin family proteins, is frequently bound at the active site [150]. There are structures available for three AraC family protein cupin superfamily domains, none of which contains a metal. Alignment of the sequences of these three AraC-family cupin superfamily domains, AraC, ToxT, and ExsA, shows that these proteins contain at most one of the four conserved His and Glu residues that comprise the metal-binding site (Fig. 53). In addition to RhaR, out of 100 sequences that we analysed, 10 additional AraC family members share the conserved metal-binding residues of the cupin motif, including RhaS (Fig. 53, and not shown). RhaS, a second rhamnose-responsive AraC family transcriptional activator protein, shares 35% sequence identity and 67% sequence similarity with RhaR [10] (sequence alignment was performed using Clustal Omega[174]). Given that all four of the metal binding residues are conserved in RhaS, it is reasonable to predict that RhaS may also contain a metal ion within its rhamnose-binding pocket. Additionally, it is likely that other AraC-family proteins with the conserved cupin superfamily metal-binding residues also bind metal ions in their cupin binding pockets.

The Ni^{2+} ion that was present in the RhaR-NTD crystal structures was likely acquired during the Ni^{2+} affinity chromatography protein purification step since the protein was exposed to Ni (but not any other divalent metal ions) during the purification. However *in vivo* it was not clear whether Ni^{2+} was the native metal for RhaR. The ionic radius of Ni^{2+} is 0.083 Å [175]; since there are a number of divalent metal ions with similar size that are physiologically relevant and commonly found in *E. coli* cells, it is possible that other metal cations with similar sizes and the same coordination number could be accommodated within the pocket *in vivo*. For instance,

Zn^{2+}, which is a six-coordination metal with ionic radius of 0.089 Å, is one possible candidate. Coincidentally, L-rhamnulose-1-phosphate aldolase which was encoded by *rhaD* binds Zn^{2+} as a

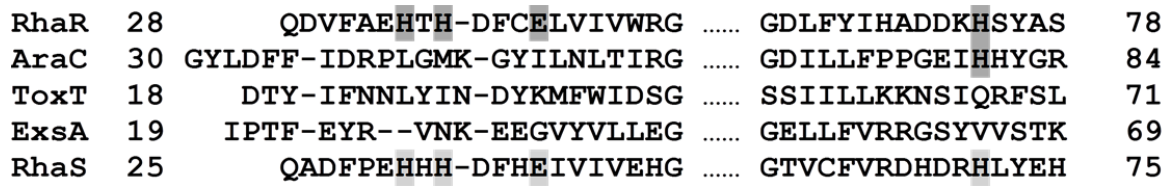


Fig. 53. Sequence alignment of RhaR, AraC, ToxT, ExsA, and ToxT. Conserved His and Glu are highlighted in gray.

cofactor at the active site. The Zn^{2+} that is bound to L-rhamnulose-1-phosphate aldolase is also coordinated by three His residues and one Glu residue. It would be very interesting if the regulator protein and the protein whose expression was regulated bind to the same metal ion. If this were true, Zn^{2+} could be a factor that affects both the transcription activation of the *rhaSR* operon and the enzymatic activity of L-rhamnulose-1-phosphate aldolase. Thus, limited access to Zn^{2+}, even when rhamnose is around, could potentailly result in *E.coli* being unable to effectively use rhamnose as a carbon source.

The comparison of the (+)rha and (-)rha RhaR-NTD structures showed that the N-terminal arm regions did not change conformation in response to rhamnose. However, two loop regions (residues V22 to T35 and residues S75 to V83) did undergo conformational changes upon rhamnose binding, suggesting that these regions may be involved in transmitting the rhamnose binding signal from the RhaR NTD to the DBD. Interestingly, the two analogous loop regions in AraC, I26 to D37 and R84 to E90, did not show any structural changes in response to arabinose. The arm region in AraC, on the other hand, upon the binding of arabinose changes its conformation. This structural change is hypothesized to release the constraint on the AraC dimer in such a way that allows the AraC dimer to bind to two adjacent half sites and activate transcription [35]. The comparison of the regions that showed structural changes in AraC and RhaR upon effector binding suggested that AraC and RhaR utilize different regions of the protein to transmit effector binding signal from one domain (NTD) to the other (DBD). Therefore, we propose that RhaR rhamnose-dependent allosteric signaling shares some features with the 'light switch' model of AraC, but differs in other features.

Model for RhaR allosteric effector response. We propose that in RhaR, there might be within-protomer changes in DNA binding affinity as well as changes in the relative positions of the

DBDs within the dimer. By comparing RhaR binding to half site and full site DNA, we could separately test the effect of rhamnose on binding by one protomer of the RhaR dimer and on binding of the full dimer, and thereby provide information about the possible mechanism. We utilized RhaR protein that was purified in the absence of rhamnose to perform *in vitro* biochemical assays, including equilibrium measurements of RhaR and RhaR-DBD binding to the full-site or half-site RhaR DNA binding sites.

One goal of our studies was to determine the effect of rhamnose on the affinity of one protomer of the RhaR dimer for its specific half-site DNA. We chose to assay this by using DNA fragments that consisted of a single half-site of the RhaR binding site, and therefore that were long enough to bind one RhaR protomer but not long enough to bind the full RhaR dimer, for several reasons. First, at the protein concentrations necessary for these experiments there was little to no monomeric wild-type RhaR detectable by gel filtration chromatography, thus isolation of monomeric wild-type RhaR wasn't possible. Second, disruption of AraC dimerization required four residue substitutions [176], and our structures showed that the RhaR dimerization interface is at least as large as that of AraC, suggesting that multiple mutations would be required to disrupt RhaR dimerization, and raising concerns that protein folding might be effected in addition to dimerization. Finally, although it would be of interest to determine whether dimerization alters the DNA binding affinity of RhaR protomers within a dimer relative to free RhaR monomers, in the event that it does, measurement of DNA binding by each protomer within a dimer was a more relevant comparison with RhaR dimer binding to its full DNA binding site for our analysis.

In addition to testing the effect of rhamnose on RhaR binding to half-site DNA, we also tested binding by the isolated RhaR-DBD to half-site DNA to identify the effect of the RhaR-

NTD on binding by each protomer. Our assays showed that deletion of the RhaR-NTD reduced DBD binding to half-site DNA by 350-fold. However, assays in the absence and presence of rhamnose indicated that rhamnose did not appreciably change the K_D for full-length RhaR binding to the half-site. Lowden *et al.* [38] found that the two HTH motifs in the DBD in the effector–bound ToxT (which is the non-activating ToxT state) were in a conformation that is quite different from those in the DNA-bound MarA, and unlikely to be able to binding DNA. They predicted that the HTH motifs of ToxT in the absence of effector would be positioned similarly to those in the DNA-bound MarA structure, thus increasing ToxT DNA binding. If HTH repositioning is important in RhaR function, the HTH repositioning would be predicted to affect the strength of DNA binding by each RhaR protomer. Our data suggests that, although HTH positioning may be important in RhaR, it does not occur in response to rhamnose binding. The 350-fold increase in half-site DNA binding by full-length RhaR relative to RhaR-DBD could potentially be a result of constitutive HTH positioning by the RhaR NTD. Thus, while effector-dependent HTH positioning may be a feature of allosteric signaling in other AraC family proteins, in RhaR this effect on DBD activity appears to occur constitutively and does not appear to be effector-dependent.

We next measured the effect of rhamnose on RhaR binding to its full DNA site and found that rhamnose increased this binding strength by 10-fold. This value is consistent with the fold-activation measured in prior *in vivo* and *in vitro* transcription activation studies [53, 173] (using chromosomally-expressed RhaR and purified RhaR, respectively), as well as prior *in vitro* measurements of DNA binding by RhaR [53]. Interestingly, the affinity of full length RhaR for binding to its full DNA site even in the absence of rhamnose is relatively strong ($K_d = 4 \times 10^{-11}$ M). If the intracellular RhaR concentration is not dramatically different than that of AraC [177,

178] (approximately 10^{-8} to 10^{-9} M), RhaR would be predicted to be bound to DNA at the *rhaSR* promoter even in the absence of rhamnose. None-the-less, and consistent with *in vitro* transcription assays (where RhaR concentration was 50-fold higher than its K_d in the absence of rhamnose [48]), there is also a 10-fold increase in transcriptional activation by RhaR upon addition of rhamnose [51]. Our previous work showed that transcriptional activation by RhaR involves protein-protein contacts with the RNA polymerase σ^{70} subunit [54]. The current findings suggest that, in the absence of rhamnose, even RhaR that is bound at the *rhaSR* promoter region is not in the optimal conformation to contact RNA polymerase, and thereby does not activate transcription. Given that DNA bending can influence transcription activation [103, 179-181], the change in DNA bending that we measured upon addition of rhamnose to RhaR could contribute to the increase in both DNA binding and transcription activation. Thus, although two independent changes could separately effect DNA binding and transcription activation, the simplest hypothesis is that the same change in RhaR structure or dynamics increases both DNA binding and transcription activation by approximately 10-fold.

Although the binding affinity of RhaR for full site DNA increased by 10-fold upon addition of rhamnose, the binding affinity of RhaR for half-site DNA did not vary upon rhamnose addition. Thus, the affinity of the dimer increased despite no increase in affinity of each protomer. The simplest hypothesis to explain these findings is that the increased affinity of the dimer involves an increase in cooperativity of DNA binding by the two DBDs. The rhamnose-dependent change in cooperativity could indicate that the RhaR DBDs are not optimally positioned to contact the two half-sites in the RhaR binding site in the absence of rhamnose, and that a rhamnose-dependent rigid body movement improves the positioning of the DBDs for cooperative binding to the full RhaR binding site. An analogous rigid body movement

is proposed in the AraC light-switch mechanism as AraC toggles from its non-activating, DNA-looped state to its activating state, bound to adjacent half-sites [35]. Further support for our hypothesis that rhamnose binding induced a rigid body movement that increased the cooperativity of binding by the two RhaR DBDs came from testing the effect of rhamnose on RhaR binding to a gapped DNA fragment. Use of the gapped DNA fragment in which the two double stranded RhaR half-sites were flexibly connected by single stranded DNA, similar to a previous assay with AraC [29], allowed us to test the prediction that the RhaR DBDs are not optimally positioned in the absence of rhamnose RhaR to bind cooperatively to the full dsDNA RhaR binding site. Addition of rhamnose increased RhaR binding to the gapped DNA by only 2-fold (compared with the 10-fold increase on fully double stranded DNA), supporting the model that the primary effect of rhamnose binding to RhaR is an increase in cooperative binding by repositioning of the DBDs in the RhaR dimer. Our finding that there is a significant difference in the angle of DNA bending by RhaR in the absence and presence of rhamnose is also consistent with a rigid body movement of the RhaR DBDs within the dimer. Thus, we hypothesize that the majority of the RhaR rhamnose allostery is due to rigid-body movement of the RhaR DBDs.

Although the mechanism of effector signaling in RhaR bears some notable similarities with that of AraC, there are also differences in the molecular details of RhaR and AraC function. One notable difference is that the isolated AraC DBD bound to half-site DNA with the same apparent equilibrium binding constant as full-length AraC in the presence of arabinose (although the DBD dissociated from half-site DNA 150-fold more slowly than AraC plus arabinose). Thus, the AraC NTD altered the kinetics of DNA binding by the DBD, but it did not alter the equilibrium binding, and therefore had a neutral effect on net DNA binding by the DBD. In

contrast, the RhaR NTD had a very large (350-fold tighter equilibrium binding constant) positive effect on equilibrium half-site DNA binding by its DBD (compare RhaR-DBD and RhaR, each binding to half-site DNA).

Another difference in the results obtained with RhaR and AraC involves the effector dependence of their DNA bend angles. Our results show a small, but significant increase of 9 degrees in the bend angle when RhaR binds to its full DNA site in the presence of rhamnose relative to the absence of rhamnose. In contrast, no change in the bend angle was detected for AraC binding to several variants of its full DNA site in the absence versus the presence of arabinose [155].

A final difference between our findings with RhaR and those with AraC involves the role of the arm region. The N-terminal arm region of AraC plays a central role in the rigid body movement of the AraC DBDs in response to binding arabinose to the NTD [35]. Unlike the structures of the AraC NTD with and without arabinose bound [27], there was no significant change in the structure of the RhaR N-terminal arm between the structures with and without rhamnose bound, suggesting that the arm may not play a significant role in transmitting the rhamnose binding signalling from the RhaR NTD to the DBD. Instead, the RhaR NTD structures show structural changes in a RhaR region that aligns with β2 of AraC and ToxT. The AraC NTD structures do not show effector-dependent structural changes in the aligned region, although molecular dynamics simulations predicted that substitutions in the AraC arm would result in the aligned AraC region differing in structure from wild type [32, 33, 182, 183]. Nonetheless, it appears that AraC and RhaR utilize different region(s) of the protein to transmit effector binding signal from the NTD to another DBD.

Overall, we propose the following model for the role of the RhaR NTD in DNA binding and the allosteric rhamnose response. Our data supports the conclusion that the RhaR NTD constitutively promotes a DBD conformation that has a much higher affinity for DNA than the DBD without the NTD. Results from other AraC family activators [38] suggests that this conformation may involve the NTD promoting a DBD conformation with the two HTH motifs in each protomer well-positioned to bind to the adjacent major grooves of the RhaR half-sites. The constitutive HTH positioning results in a much higher basal level of transcription activation than RhaR would have if HTH positioning was rhamnose dependent. The high basal level would allow RhaR to quickly activate transcription of the *rhaSR* operon as soon as rhamnose became available. The constitutive HTH positioning fits the role of RhaR. Prior [51, 53] and current results indicate that the addition of rhamnose increases RhaR DNA binding, transcription activation (even when DNA binding is saturated) and DNA bending. We propose that rhamnose binding to RhaR NTD induces a rigid body movement of the RhaR DBDs relative to the NTDs that increases cooperative binding to the two half-sites of the full RhaR binding site; likely explaining the increases in DNA binding and bending. We hypothesize that the same rigid body movement may also increase transcription activation by increasing RhaR contacts with RNAP and/or DNA bending.

Allosteric control of transcription activation is common in the AraC family, therefore, investigating how effector binding impacts protein conformation at an atomic level may provide us with insight into the functions and mechanisms of allosteric control in the broader AraC family [4]. Our data indicates that rigid body movement is likely involved in RhaR rhamnose allosteric signalling mechanism. In RhaS, our data currently suggests that HTH positioning is involved in the mechanism by which RhaS responds to rhamnose signalling. In AraC, rigid body

movement seems to be the major conformational change the protein undergoes in respond to arabinose. We thus further propose that the rigid body movement and HTH positioning mechanisms are somewhat common as mechanisms to respond to effector molecules among AraC family proteins. Additionally, the data suggesting that HTH positioning occurs in response to the presence of RhaR-NTD could potentially explain why many virulence activators within this family require the presence of NTD to activate transcription but do not dimerize or need an effector. This study is a step forward in both the characterization of RhaR and in attempting to develop a family-wide model of the mechanism of allosteric regulation. As many AraC family members are regulated by allosteric interactions, including some that mediate virulence factor expression in pathogenic bacteria, elucidating a family-wide mechanism that they use to respond to allosteric interactions would represent significant progress towards characterizing hundreds of currently uninvestigated proteins, as well as providing potential targets for novel antimicrobial compounds designed to stop, stimulate, or imitate allosteric signalling in AraC family proteins.

REFERENCES

1. Ibarra, J.A., et al., *The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors.* Genetica, 2008. **133**(1): p. 65-76.
2. Gallegos, M.T., C. Michan, and J.L. Ramos, *The XylS/AraC family of regulators.* Nucleic Acids Res, 1993. **21**(4): p. 807-10.
3. Egan, S.M., *Growing repertoire of AraC/XylS activators.* J Bacteriol, 2002. **184**(20): p. 5529-32.
4. Gallegos, M.T., et al., *Arac/XylS family of transcriptional regulators.* Microbiol Mol Biol Rev, 1997. **61**(4): p. 393-410.
5. J. F. Tobin, R.F.S., *Positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes.* 1987.
6. Ramos, J.L., et al., *A family of positive regulators related to the Pseudomonas putida TOL plasmid XylS and the Escherichia coli AraC activators.* Nucleic Acids Res, 1990. **18**(8): p. 2149-52.
7. Kwon, H.J., et al., *Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA.* Nat Struct Biol, 2000. **7**(5): p. 424-30.
8. Rhee, S., et al., *A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10413-8.
9. Tobes, R. and J.L. Ramos, *AraC-XylS database: a family of positive transcriptional regulators in bacteria.* Nucleic Acids Research, 2002. **30**(1): p. 318-321.
10. Egan, S.M. and R.F. Schleif, *A regulatory cascade in the induction of rhaBAD.* J Mol Biol, 1993. **234**(1): p. 87-98.

11. Englesberg, E., *[3] Isolation of mutants in the l-arabinose gene-enzyme complex*, in *Methods in Enzymology*. 1966, Academic Press. p. 15-21.

12. Spooner, R.A., K. Lindsay, and F.C. Franklin, *Genetic, functional and sequence analysis of the xylR and xylS regulatory genes of the TOL plasmid pWW0.* J Gen Microbiol, 1986. **132**(5): p. 1347-58.

13. Webster, C., L. Gardner, and S. Busby, *The Escherichia coli melR gene encodes a DNA-binding protein with affinity for specific sequences located in the melibiose-operon regulatory region.* Gene, 1989. **83**(2): p. 207-13.

14. Demple, B., et al., *Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis.* Proc Natl Acad Sci U S A, 1985. **82**(9): p. 2688-92.

15. Hakura, A., et al., *Cloning and characterization of the Salmonella typhimurium ada gene, which encodes O6-methylguanine-DNA methyltransferase.* J Bacteriol, 1991. **173**(12): p. 3663-72.

16. Morohoshi, F., K. Hayashi, and N. Munakata, *Bacillus subtilis ada operon encodes two DNA alkyltransferases.* Nucleic Acids Res, 1990. **18**(18): p. 5473-80.

17. Amabile-Cuevas, C.F. and B. Demple, *Molecular characterization of the soxRS genes of Escherichia coli: two genes control a superoxide stress regulon.* Nucleic Acids Res, 1991. **19**(16): p. 4479-84.

18. Cohen, S.P., H. Hachler, and S.B. Levy, *Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli.* J Bacteriol, 1993. **175**(5): p. 1484-92.

19. Champion, G.A., et al., *A branch in the ToxR regulatory cascade of Vibrio cholerae revealed by characterization of toxT mutant strains.* Mol Microbiol, 1997. **23**(2): p. 323-31.

20. DiRita, V.J., et al., *Regulatory cascade controls virulence in Vibrio cholerae.* Proc Natl Acad Sci U S A, 1991. **88**(12): p. 5403-7.

21. Higgins, D.E., E. Nazareno, and V.J. DiRita, *The virulence gene activator ToxT from Vibrio cholerae is a member of the AraC family of transcriptional activators.* J Bacteriol, 1992. **174**(21): p. 6974-80.

22. Yu, R.R. and V.J. DiRita, *Analysis of an autoregulatory loop controlling ToxT, cholera toxin, and toxin-coregulated pilus production in Vibrio cholerae.* J Bacteriol, 1999. **181**(8): p. 2584-92.

23. Brown, R.C. and R.K. Taylor, *Organization of tcp, acf, and toxT genes within a ToxT-dependent operon.* Mol Microbiol, 1995. **16**(3): p. 425-39.

24. Caron, J., L.M. Coffield, and J.R. Scott, *A plasmid-encoded regulatory gene, rns, required for expression of the CS1 and CS2 adhesins of enterotoxigenic Escherichia coli.* Proc Natl Acad Sci U S A, 1989. **86**(3): p. 963-7.

25. Adler, B., et al., *A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of Shigella flexneri.* Mol Microbiol, 1989. **3**(5): p. 627-35.

26. Hovey, A.K. and D.W. Frank, *Analyses of the DNA-binding and transcriptional activation properties of ExsA, the transcriptional activator of the Pseudomonas aeruginosa exoenzyme S regulon.* Journal of bacteriology, 1995. **177**(15): p. 4427-36.

27. Saviola, B., R. Seabold, and R.F. Schleif, *Arm-domain interactions in AraC.* J Mol Biol, 1998. **278**(3): p. 539-48.

28. Reed, W.L. and R.F. Schleif, *Hemiplegic mutations in AraC protein.* J Mol Biol, 1999. **294**(2): p. 417-25.

29. Harmer, T., M. Wu, and R. Schleif, *The role of rigidity in DNA looping-unlooping by AraC.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 427-31.

30. Wu, M. and R. Schleif, *Strengthened arm-dimerization domain interactions in AraC.* J Biol Chem, 2001. **276**(4): p. 2562-4.

31. Wu, M. and R. Schleif, *Mapping arm-DNA-binding domain interactions in AraC.* J Mol Biol, 2001. **307**(4): p. 1001-9.

32. Weldon, J.E., et al., *Structure and properties of a truely apo form of AraC dimerization domain.* Proteins, 2007. **66**(3): p. 646-54.

33. Soisson, S.M., et al., *Structural basis for ligand-regulated oligomerization of AraC.* Science, 1997. **276**(5311): p. 421-425.

34. Rodgers, M.E. and R. Schleif, *Solution structure of the DNA binding domain of AraC protein.* Proteins, 2009. **77**(1): p. 202-8.

35. Schleif, R., *AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action.* FEMS Microbiology Reviews, 2010. **34**(5): p. 779-796.

36. Malaga, F., et al., *A genetic and physical study of the interdomain linker of E. Coli AraC protein-a trans-subunit communication pathway.* Proteins, 2016. **84**(4): p. 448-60.

37. Schleif, R., *Regulation of the L-arabinose operon of Escherichia coli.* Trends Genet, 2000. **16**(12): p. 559-65.

38. Lowden, M.J., et al., *Structure of Vibrio cholerae ToxT reveals a mechanism for fatty acid regulation of virulence genes.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2860-5.

39. Schuhmacher, D.A. and K.E. Klose, *Environmental signals modulate ToxT-dependent virulence factor expression in Vibrio cholerae.* J Bacteriol, 1999. **181**(5): p. 1508-14.

40. Chatterjee, A., P.K. Dutta, and R. Chowdhury, *Effect of fatty acids and cholesterol present in bile on expression of virulence factors and motility of Vibrio cholerae.* Infect Immun, 2007. **75**(4): p. 1946-53.

41. Childers, B.M., et al., *N-terminal residues of the Vibrio cholerae virulence regulatory protein ToxT involved in dimerization and modulation by fatty acids.* J Biol Chem, 2011. **286**(32): p. 28644-55.

42. Cole, S.D. and R. Schleif, *A new and unexpected domain-domain interaction in the AraC protein.* Proteins, 2012. **80**(5): p. 1465-75.

43. Baldoma, L., et al., *Cloning, mapping and gene product identification of rhaT from Escherichia coli K12.* FEMS Microbiol Lett, 1990. **60**(1-2): p. 103-7.

44. Power, J., *The L-rhamnose genetic system in Escherichia coli K-12.* Genetics, 1967. **55**(3): p. 557-68.

45. Tate, C.G., J.A. Muiry, and P.J. Henderson, *Mapping, cloning, expression, and sequencing of the rhaT gene, which encodes a novel L-rhamnose-H^+ transport protein in Salmonella typhimurium and Escherichia coli.* J Biol Chem, 1992. **267**(10): p. 6923-32.

46. Tobin, J.F. and R.F. Schleif, *Positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes.* J Mol Biol, 1987. **196**(4): p. 789-99.

47. Badia, J., et al., *Identification of the rhaA, rhaB and rhaD gene products from Escherichia coli K-12.* FEMS Microbiol Lett, 1989. **53**(3): p. 253-7.

48. Tobin, J.F. and R.F. Schleif, *Transcription from the rha operon psr promoter.* J Mol Biol, 1990. **211**(1): p. 1-4.

49. Via, P., et al., *Transcriptional regulation of the Escherichia coli rhaT gene.* Microbiology, 1996. **142 (Pt 7)**: p. 1833-40.

50. Egan, S.M. and R.F. Schleif, *DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at rhaBAD.* J Mol Biol, 1994. **243**(5): p. 821-9.

51. Holcroft, C.C. and S.M. Egan, *Interdependence of activation at rhaSR by cyclic AMP receptor protein, the RNA polymerase alpha subunit C-terminal domain, and rhaR.* J Bacteriol, 2000. **182**(23): p. 6774-82.

52. Kolin, A., et al., *Differences in the mechanism of the allosteric L-rhamnose responses of the AraC/XylS family transcription activators RhaS and RhaR.* Mol Microbiol, 2008. **68**(2): p. 448-61.

53. Tobin, J.F. and R.F. Schleif, *Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of Escherichia coli.* J Mol Biol, 1990. **211**(1): p. 75-89.

54. Wickstrum, J.R. and S.M. Egan, *Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR.* J Bacteriol, 2004. **186**(18): p. 6277-85.

55. Wickstrum, J.R., et al., *Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain.* J Bacteriol, 2007. **189**(14): p. 4984-93.

56. Bhende, P.M. and S.M. Egan, *Genetic evidence that transcription activation by RhaS involves specific amino acid contacts with sigma 70.* J Bacteriol, 2000. **182**(17): p. 4959-69.

57. Kolin, A., et al., *Linker regions of the RhaS and RhaR proteins.* J Bacteriol, 2007. **189**(1): p. 269-71.

58. Shrestha, M., et al., *Structural Analysis of the Regulatory Domain of ExsA, a Key Transcriptional Regulator of the Type Three Secretion System in Pseudomonas aeruginosa.* PLoS One, 2015. **10**(8): p. e0136533.

59. Jennison, A.V. and N.K. Verma, *Shigella flexneri infection: pathogenesis and vaccine development.* FEMS Microbiol Rev, 2004. **28**(1): p. 43-58.

60. Kotloff, K.L., et al., *Global burden of Shigella infections: implications for vaccine development and implementation of control strategies.* Bulletin of the World Health Organization, 1999. **77**(8): p. 651-66.

61. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* Lancet., 2013. **382**(9888): p. 209-222.

62. Sansonetti, P.J., D.J. Kopecko, and S.B. Formal, *Involvement of a plasmid in the invasive ability of Shigella flexneri.* Infect Immun, 1982. **35**(3): p. 852-60.

63. Blocker, A., et al., *Structure and composition of the Shigella flexneri "needle complex", a part of its type III secreton.* Mol Microbiol, 2001. **39**(3): p. 652-63.

64. High, N., et al., *IpaB of Shigella flexneri causes entry into epithelial cells and escape from the phagocytic vacuole.* EMBO J, 1992. **11**(5): p. 1991-9.

65. Zychlinsky, A., et al., *IpaB mediates macrophage apoptosis induced by Shigella flexneri.* Mol Microbiol, 1994. **11**(4): p. 619-27.

66. Sasakawa, C., et al., *Eight genes in region 5 that form an operon are essential for invasion of epithelial cells by Shigella flexneri 2a.* J Bacteriol, 1993. **175**(8): p. 2334-46.

67. Murphy, E.R. and S.M. Payne, *RyhB, an iron-responsive small RNA molecule, regulates Shigella dysenteriae virulence.* Infect Immun, 2007. **75**(7): p. 3470-7.

68. Gore, A.L. and S.M. Payne, *CsrA and Cra influence Shigella flexneri pathogenesis.* Infect Immun, 2010. **78**(11): p. 4674-82.

69. Tobe, T., et al., *Temperature-regulated expression of invasion genes in Shigella flexneri is controlled through the transcriptional activation of the virB gene on the large plasmid.* Mol Microbiol, 1991. **5**(4): p. 887-93.

70. Allaoui, A., et al., *icsB: a Shigella flexneri virulence gene necessary for the lysis of protrusions during intercellular spread.* Mol Microbiol, 1992. **6**(12): p. 1605-16.

71. Le Gall, T., et al., *Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of Shigella flexneri.* Microbiology, 2005. **151**(Pt 3): p. 951-62.

72. Goldberg, M.B., J.A. Theriot, and P.J. Sansonetti, *Regulation of surface presentation of IcsA, a Shigella protein essential to intracellular movement and spread, is growth phase dependent.* Infect Immun, 1994. **62**(12): p. 5664-8.

73. Brotcke Zumsteg, A., et al., *IcsA is a Shigella flexneri adhesin regulated by the type III secretion system and required for pathogenesis.* Cell Host Microbe, 2014. **15**(4): p. 435-45.

74. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2013*. 2013; Available from: http://www.cdc.gov/drugresistance/threat-report-2013/.

75. Clatworthy, A.E., E. Pierson, and D.T. Hung, *Targeting virulence: a new paradigm for antimicrobial therapy.* Nat Chem Biol, 2007. **3**(9): p. 541-8.

76. Bowser, T.E., et al., *Novel anti-infection agents: small-molecule inhibitors of bacterial transcription factors.* Bioorg Med Chem Lett, 2007. **17**(20): p. 5652-5.

77. Kim, O.K., et al., *N-hydroxybenzimidazole inhibitors of the transcription factor LcrF in Yersinia: novel antivirulence agents.* Journal of medicinal chemistry, 2009. **52**(18): p. 5626-34.

78. Garrity-Ryan, L.K., et al., *Small molecule inhibitors of LcrF, a Yersinia pseudotuberculosis transcription factor, attenuate virulence and limit infection in a murine pneumonia model.* Infection and immunity, 2010. **78**(11): p. 4683-90.

79. Hung, D.T., et al., *Small-molecule inhibitor of Vibrio cholerae virulence and intestinal colonization.* Science, 2005. **310**(5748): p. 670-4.

80. Koppolu, V., et al., *Small-molecule inhibitor of the Shigella flexneri master virulence regulator VirF.* Infect Immun, 2013. **81**(11): p. 4220-31.

81. Skredenske, J.M., et al., *Identification of a small-molecule inhibitor of bacterial AraC family activators.* J Biomol Screen, 2013. **18**(5): p. 588-98.

82. Emanuele, A.A. and G.A. Garcia, *Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the Shigella flexneri Virulence Regulator VirF.* PLoS One, 2015. **10**(9): p. e0137410.

83. Grier, M.C., et al., *N-Hydroxybenzimidazole inhibitors of ExsA MAR transcription factor in Pseudomonas aeruginosa: In vitro anti-virulence activity and metabolic stability.* Bioorganic & medicinal chemistry letters, 2010. **20**(11): p. 3380-3.

84. Hurt, J.K., et al., *High-throughput screening of the virulence regulator VirF: a novel antibacterial target for shigellosis.* Journal of biomolecular screening, 2010. **15**(4): p. 379-87.

85. Backman, K., Y.M. Chen, and B. Magasanik, *Physical and genetic characterization of the glnA--glnG region of the Escherichia coli chromosome.* Proc Natl Acad Sci U S A, 1981. **78**(6): p. 3743-7.

86. Stewart, G.S., et al., *pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression.* Plasmid, 1986. **15**(3): p. 172-81.

87. Bhende, P.M. and S.M. Egan, *Amino acid-DNA contacts by RhaS: an AraC family transcription activator.* J Bacteriol, 1999. **181**(17): p. 5185-92.

88. Neidhardt, F.C., P.L. Bloch, and D.F. Smith, *Culture medium for enterobacteria.* J Bacteriol, 1974. **119**(3): p. 736-47.

89. Miller, J.H., *Experiments in molecular genetics*. 1972: Cold Spring Harbor Laboratory.

90. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.

91. Kabsch, W., *Automatic indexing of rotation diffraction patterns.* Journal of Applied Crystallography, 1988. **21**(1): p. 67-72.

92. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 282-92.

93. Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.

94. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

95. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

96. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

97. Painter, J. and E.A. Merritt, *Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 4): p. 439-50.

98. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.

99. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

100. Potterton, L., et al., *Developments in the CCP4 molecular-graphics project.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2288-94.

101. Otwinowski, Z. and W. Minor, *[20] Processing of X-ray diffraction data collected in oscillation mode*, in *Methods in Enzymology*. 1997, Academic Press. p. 307-326.

102. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallographica Section D, 1999. **55**(4): p. 849-861.

103. Wu, H.M. and D.M. Crothers, *The locus of sequence-directed and protein-induced DNA bending.* Nature, 1984. **308**(5959): p. 509-13.

104. Thompson, J.F. and A. Landy, *Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes.* Nucleic Acids Res, 1988. **16**(20): p. 9687-705.

105. Student, *The Probable Error of a Mean.* Biometrika, 1908. **6**(1): p. 1-25.

106. Rippe, K., *Analysis of protein-DNA binding at equilibrium.* B.I.F *Futura*, 1997. **12**: p. 20-26.

107. Perez-Howard, G.M., P.A. Weil, and J.M. Beechem, *Yeast TATA binding protein interaction with DNA: fluorescence determination of oligomeric state, equilibrium binding, on-rate, and dissociation kinetics.* Biochemistry, 1995. **34**(25): p. 8005-17.

108. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery.* Journal of biomolecular screening, 2001. **6**(6): p. 429-40.

109. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, 2006. **357**(2): p. 289-98.

110. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 3rd ed. 2006, New York: Springer. xxvi, 954 p.

111. Rankin, C.A., et al., *Isolated RING2 domain of parkin is sufficient for E2-dependent E3 ligase activity.* Biochemistry, 2014. **53**(1): p. 225-34.

112. Koenig, T., et al., *Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics.* J Proteome Res, 2008. **7**(9): p. 3708-17.

113. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

114. Grosdidier, A., V. Zoete, and O. Michielin, *SwissDock, a protein-small molecule docking web service based on EADock DSS.* Nucleic acids research, 2011. **39**(Web Server issue): p. W270-7.

115. Lee, H.S. and Y. Zhang, *BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures.* Proteins, 2012. **80**(1): p. 93-110.

116. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.

117. Chi, G., et al., *Implications of binding mode and active site flexibility for inhibitor potency against the salicylate synthase from Mycobacterium tuberculosis.* Biochemistry, 2012. **51**(24): p. 4868-79.

118. Hau, J.C., et al., *Leveraging the contribution of thermodynamics in drug discovery with the help of fluorescence-based thermal shift assays.* J Biomol Screen, 2011. **16**(5): p. 552-6.

119. Lo, M.C., et al., *Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery.* Anal Biochem, 2004. **332**(1): p. 153-9.

120. Ross, J.J., U. Gryczynski, and R. Schleif, *Mutational analysis of residue roles in AraC function.* J Mol Biol, 2003. **328**(1): p. 85-93.

121. Rosenfeld, L., et al., *Combinatorial and Computational Approaches to Identify Interactions of Macrophage Colony-stimulating Factor (M-CSF) and Its Receptor c-FMS.* J Biol Chem, 2015. **290**(43): p. 26180-93.

122. Evans, P., *Scaling and assessment of data quality.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.

123. Diederichs, K. and P.A. Karplus, *Improved R-factors for diffraction data analysis in macromolecular crystallography.* Nat Struct Biol, 1997. **4**(4): p. 269-75.

124. Weiss, M.S., *Global indicators of X-ray data quality.* Journal of Applied Crystallography, 2001. **34**: p. 130-135.

125. Karplus, P.A. and K. Diederichs, *Linking crystallographic model and data quality.* Science, 2012. **336**(6084): p. 1030-3.

126. Evans, P., *Biochemistry. Resolving some old problems in protein crystallography.* Science, 2012. **336**(6084): p. 986-7.

127. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.

128. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.

129. Urbanowski, M.L., G.L. Lykken, and T.L. Yahr, *A secreted regulatory protein couples transcription to the secretory activity of the Pseudomonas aeruginosa type III secretion system.* Proc Natl Acad Sci U S A, 2005. **102**(28): p. 9930-5.

130. Jain, P., et al., *3-Substituted Biquinolinium Inhibitors of AraC Family Transcriptional Activator VirF from Obtained Through In Situ Chemical Ionization of 3,4-Disubstituted Dihydroquinolines.* RSC Adv, 2014. **4**(75): p. 39809-39816.

131. Pitten, F.A. and A. Kramer, *Efficacy of cetylpyridinium chloride used as oropharyngeal antiseptic.* Arzneimittelforschung, 2001. **51**(7): p. 588-95.

132. Scheie, A., *Models of action of currently known chemical antiplaque agents other than chlorhexidine.* Journal of dental research 1989. **68**: p. 1609-1616.

133. *Electrophile*. Available from: https://en.wikipedia.org/wiki/Electrophile.

134. Mingeot-Leclercq, M.P., Y. Glupczynski, and P.M. Tulkens, *Aminoglycosides: activity and resistance.* Antimicrob Agents Chemother, 1999. **43**(4): p. 727-37.

135. Taber, H.W., et al., *Bacterial uptake of aminoglycoside antibiotics.* Microbiol Rev, 1987. **51**(4): p. 439-57.

136. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.

137. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-74.

138. Dunwell, J.M., S. Khuri, and P.J. Gane, *Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily.* Microbiol Mol Biol Rev, 2000. **64**(1): p. 153-79.

139. Dundas, J., et al., *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W116-8.

140. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

141. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.

142. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

143. Seabold, R.R. and R.F. Schleif, *Apo-AraC actively seeks to loop.* J Mol Biol, 1998. **278**(3): p. 529-38.

144. Ghosh, M. and R.F. Schleif, *Biophysical evidence of arm-domain interactions in AraC.* Anal Biochem, 2001. **295**(1): p. 107-12.

145. Laskowski, R.A., et al., *PDBsum: a Web-based database of summaries and analyses of all PDB structures.* Trends Biochem Sci, 1997. **22**(12): p. 488-90.

146. Laskowski, R.A., *PDBsum: summaries and analyses of PDB structures.* Nucleic Acids Res, 2001. **29**(1): p. 221-2.

147. Laskowski, R.A., V.V. Chistyakov, and J.M. Thornton, *PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids.* Nucleic Acids Res, 2005. **33**(Database issue): p. D266-8.

148. Laskowski, R.A., *PDBsum new things.* Nucleic Acids Res, 2009. **37**(Database issue): p. D355-9.

149. Soisson, S.M., et al., *The 1.6 A crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose.* J Mol Biol, 1997. **273**(1): p. 226-37.

150. Dunwell, J.M., A. Purvis, and S. Khuri, *Cupins: the most functionally diverse protein superfamily?* Phytochemistry, 2004. **65**(1): p. 7-17.

151. Miller, J.H., *Experiments in Molecular Genetics*. 1972, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. 433.

152. Krissinel, E., *Enhanced fold recognition using efficient short fragment clustering.* Journal of Molecular Biochemistry, 2012. **1**(2): p. 76-85.

153. Harding, M.M., *Geometry of metal-ligand interactions in proteins.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 3): p. 401-11.

154. Zheng, H., et al., *Data mining of metal ion environments present in protein structures.* J Inorg Biochem, 2008. **102**(9): p. 1765-76.

155. Saviola, B., R.R. Seabold, and R.F. Schleif, *DNA bending by AraC: a negative mutant.* J Bacteriol, 1998. **180**(16): p. 4227-32.

156. Wickstrum, J.R., et al., *The AraC/XylS Family Activator RhaS Negatively Autoregulates rhaSR Expression by Preventing CRP from Contributing to Activation.* J Bacteriol, 2009. **192**(1): p. 225-232.

157. Chernov, A.A., *Protein crystals and their growth.* J Struct Biol, 2003. **142**(1): p. 3-21.

158. Newman, J., J. Xu, and M.C. Willis, *Initial evaluations of the reproducibility of vapor-diffusion crystallization.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 7): p. 826-32.

159. Yin DC, W.N.I., Lu HM, Ye YJ, Li HS, Luo HM, Inatomi Y., *Uncertainties in crystallization of hen-egg white lysozyme: reproducibility issue.* Cryst. Res. Technol., 2008. **43**(No. 4): p. 447 – 454.

160. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization.* Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 1): p. 2-20.

161. Asherie, N., *Protein crystallization and phase diagrams.* Methods, 2004. **34**(3): p. 266-72.

162. Lowden, M.J., Kull, F.J. *Crystal Structure of ToxT from Vibrio Cholerae O395*. 2010; Available from: http://www.rcsb.org/pdb/explore.do?structureId=3GBG.

163. Lovell, S., Wehmeyer, G., Battaile, K.P., Li, J., Egan, S. *1.65A resolution structure of ToxT from Vibrio cholerae (P21 Form)*. 2016; Available from: http://www.rcsb.org/pdb/explore/explore.do?structureId=4mlo.

164. Childers, B.M., et al., *Identification of residues critical for the function of the Vibrio cholerae virulence regulator ToxT by scanning alanine mutagenesis.* J Mol Biol, 2007. **367**(5): p. 1413-30.

165. Shakhnovich, E.A., et al., *Virstatin inhibits dimerization of the transcriptional activator ToxT.* Proc Natl Acad Sci U S A, 2007. **104**(7): p. 2372-7.

166. Deprez-Poulain, R. and B. Deprez, *Facts, figures and trends in lead generation.* Curr Top Med Chem, 2004. **4**(6): p. 569-80.

167. Fruber M, N.F., Steele J, *Lead Generation*, in *Handbook of Medicinal Chemistry: Principles and Practice.*, W.S. Davis A, Editor. 2013: Royal Society of Chemistry. p. 505–528.

168. Keseru, G.M. and G.M. Makara, *Hit discovery and hit-to-lead approaches.* Drug Discov Today, 2006. **11**(15-16): p. 741-8.

169. Bleicher, K.H., et al., *Hit and lead generation: beyond high-throughput screening.*

170. Renton-Harper, P., et al., *A comparison of chlorhexidine, cetylpyridinium chloride, triclosan, and C31G mouthrinse products for plaque inhibition.* J Periodontol, 1996. **67**(5): p. 486-9.

171. Bodor, N. and P. Buchwald, *Soft drug design: general principles and recent applications.* Med Res Rev, 2000. **20**(1): p. 58-101.

172. Bodor, N., J.J. Kaminski, and S. Selk, *Soft drugs. 1. Labile quaternary ammonium salts as soft antimicrobials.* J Med Chem, 1980. **23**(5): p. 469-74.

173. Wickstrum, J.R., T.J. Santangelo, and S.M. Egan, *Cyclic AMP receptor protein and RhaR synergistically activate transcription from the L-rhamnose-responsive rhaSR promoter in Escherichia coli.* J Bacteriol, 2005. **187**(19): p. 6708-18.

174. Sievers, F., et al., *Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega.* Molecular Systems Biology, 2011. **7**(1).

175. Shannon, R., *Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides.* Acta Crystallographica Section A, 1976. **32**(5): p. 751-767.

176. LaRonde-LeBlanc, N. and C. Wolberger, *Characterization of the oligomeric states of wild type and mutant AraC.* Biochemistry, 2000. **39**(38): p. 11593-601.

177. Hahn, S. and R. Schleif, *In vivo regulation of the Escherichia coli araC promoter.* J Bacteriol, 1983. **155**(2): p. 593-600.

178. Kolodrubetz, D. and R. Schleif, *Identification of araC protein and two-dimensional gels, its in vivo instability and normal level.* J Mol Biol, 1981. **149**(1): p. 133-9.

179. Perez-Martin, J. and M. Espinosa, *Protein-induced bending as a transcriptional switch.* Science, 1993. **260**: p. 805-807.

180. Perez-Martin, J., F. Rojo, and V. de Lorenzo, *Promoters responsive to DNA bending: a common theme in prokaryotic gene expression.* Microbiol Rev, 1994. **58**(2): p. 268-90.

181. Ohyama, T., *DNA conformation and transcription*. Molecular biology intelligence unit. 2005, Georgetown, Tex. New York, NY.: Landes Bioscience ;Springer Science Business Media. 211 p.

182. Berrondo, M., J.J. Gray, and R. Schleif, *Computational predictions of the mutant behavior of AraC.* J Mol Biol, 2010. **398**(3): p. 462-70.

183. Dirla, S., J.Y. Chien, and R. Schleif, *Constitutive mutations in the Escherichia coli AraC protein.* J Bacteriol, 2009. **191**(8): p. 2668-74.